CHARACTERIZATION OF THE MOUSE PEA3 PROMOTER

MOLECULAR AND FUNCTIONAL CHARACTERIZATION OF THE MOUSE PEA3 PROMOTER

By

JANE MARIE BARRETT, B.Sc.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Master of Science

McMaster University

© Copyright by Jane Marie Barrett, July 1997

MASTER OF SCIENCE (1997)
(Biochemistry)

McMaster University
Hamilton, Ontario

TITLE: Molecular and Functional Characterization of the Mouse PEA3 Promoter

AUTHOR: Jane Marie Barrett, B.Sc. (Queen's University)

SUPERVISOR: Professor J.A. Hassell

NUMBER OF PAGES: xiv, 120

ABSTRACT

PEA3 is a member of the expanding Ets family of transcription factors. In the adult mouse, PEA3 mRNA is expressed at highest levels in the brain, epididymis and at lower levels in the mammary gland, testes, ovary and uterus. PEA3 mRNA is expressed differentially during mouse embryogenesis and is down-regulated following retinoic acid induced differentiation in mouse embryonal carcinoma cell lines. PEA3 is overexpressed at the transcriptional level in 93% of all HER2/neu positive human breast tumors. The molecular basis for differential transcription of the PEA3 gene is not known. Sequence analysis revealed that the upstream region of the PEA3 gene has characteristics of a CpG island and does not possess a recognizable "TATA" element. Rapid amplification of 5' cDNA ends (5'RACE) reveals that transcription initiates from multiple sites, consistent with the absence of TATA elements. To localize *cis*-acting sequences required for PEA3 expression, deletions of the putative promoter were placed upstream of a luciferase reporter gene and tested for activity in the FM3A cell line. FM3A cells express substantial levels of PEA3 mRNA and protein, which suggests that all of the factors required for transcription are present in the cells. Transient transfections of 5' and 3' deletion mutants of the PEA3 promoter indicated that the efficiency of the PEA3 promoter depended on both negative and positive *cis*-elements, located upstream and downstream of the transcription start sites. A DNA fragment containing a region from -3 to +676, relative to the major start site of transcription, was sufficient for maximal promoter activity.

Luciferase reporter plasmids containing more 5' flanking sequence had lower activity indicating the presence of silencer elements. To aid the identification of critical sequence elements within the minimal PEA3 promoter, we cloned and sequenced the putative human PEA3 promoter. Comparison of the mouse and human PEA3 DNAs revealed that sequences required for maximal promoter activity in the mouse were highly conserved in the human gene. Furthermore, these conserved sequences corresponded to a variety of consensus binding sites: 6 Sp1, 8 c-ets-1, 3 PEA3, 3 AP-2, 3 MZF-1, 2 MyoD, 2 Ik-1, 2 c/EBPβ, 2 δEF-1/USF, 2 HSF1 and one of each of the following: AP-4, Ik-2, SRY, CP2, HEN-1, CREB and E47.

ACKNOWLEDGMENTS

I wish to thank my supervisor, Dr. John A. Hassell, for his assistance and guidance throughout my studies. It was a pleasure to work for a someone who listened to my ideas and offered fair criticism. I would also like to thank my parents for their support of my efforts and especially to my husband, Joe; he is an excellent listener and my best friend.

CONTRIBUTIONS FROM OTHERS

I gratefully acknowledge the contributions from others for the work presented in this thesis. Oligonucleotide primers were prepared by Dinsdale Gooden and all sequencing reactions were performed by Brian Allore in the Central Facility at the Institute for Molecular Biology and Biotechnology, McMaster University. The fluorescence *in situ* hybridization (FISH) analyses were performed by Dr. Barbara Beatty and members of the Human Genome FISH Mapping Resource Centre at the Hospital for Sick Children, Toronto.

TABLE OF CONTENTS

Abstract iii

Acknowledgments v

Contributions vi

Table of Contents vii

List of Tables x

List of Figures xi

List of Appendices xiii

Abbreviations xiv

Introduction 1

1. PEA3 is a member of the *Ets* oncogene family 1
2. Ets proteins and oncogenesis 3
3. Initiation of transcription by RNA Polymerase II 8
4. The Luciferase Expression System 10
5. Characterization of the Mouse PEA3 Gene Promoter 14

Objectives 22

Materials 23

Methods 24

1. Construction of Mouse PEA3/pGL3 Chimera 24
   a. PCR Amplification of the Putative Promoter Region of Mouse PEA3 24
   b. Restriction Endonuclease Isolation of the Putative Promoter Region of PEA3 25
   c. Purification and Modification of PCR Amplified DNA 25

d. Preparation of Vector DNA 26
e. Ligation Reaction 27
f. Transformation 27
2. Identification and Characterization of Positive Chimera 28
a. Restriction Endonuclease Analysis of Positive Clones 28
b. DNA Sequencing with GL2 and RV3 Sequencing Primers 28
3. Luciferase Reporter Assay 29
a. Large Scale Preparation of Luciferase Constructs 29
b. Transfection of FM3A cells 29
c. Preparation of Cell Extracts 30
d. Luciferase and β-Galactosidase Assays 30
4. Statistical Analysis 31
5. Isolation and Characterization of Human PEA3 Cosmids 32
a. Human Cosmid Library Screening 32
b. Restriction Endonuclease and Southern Blot Analysis of Human PEA3 Cosmids 34
c. Subcloning of DNA Fragments from the Human PEA3 locus 34
d. Chromosomal Mapping of PEA3 to Human Chromosomes by FISH 35
e. Sequencing of the Human PEA3 Locus 35
f. Computer-Assisted Sequence Analysis and Alignment Studies of Human and Mouse PEA3 35

Results 37

1. PCR Amplification of the Putative Promoter Region for PEA3 38
a. Restriction Endonuclease and Sequencing Analysis 42
2. Luciferase Assays 48
a. Promoter Efficiency of 5' Deletions of the Putative Promoter 48
b. Promoter Efficiency of 3' Deletions of the Putative Promoter 52
c. Assessment of SV40 Enhancer-Containing Luciferase Constructs 55
3. Isolation and Characterization of Human PEA3 Cosmids 58
a. Human Cosmid Library Screening 58
b. Restriction Endonuclease and Southern Blot Analysis of Human PEA3 Cosmids 61
c. Subcloning of DNA Fragments From the Human PEA3 Locus 66
d. Chromosomal Mapping of PEA3 to Human Chromosomes by FISH 69
e. Sequencing of the Human PEA3 Locus 69
f. Computer-Assisted Sequence Analysis and Alignment Studies of Human and Mouse PEA3 74

Discussion 86

1. Cloning and Detection in the pGL3 Vector System 86
2. Luciferase Assays 87
    a. General Analysis of Promoter Activity 88
    b. 5' Deletions of the Putative Promoter 89
    c. 3' Deletions of the Putative Promoter 91
    d. The Effect of the SV40 Enhancer on the PEA3 Promoter 96
    e. Factors Regulating PEA3 Expression 97
3. Isolation and Sequencing of the Human PEA3 Locus 102
4. Chromosomal Localization of PEA3 103

Conclusions 105

References 106

LIST OF TABLES

Table 1: mRNA Expression of PEA3 Subfamily Members 4

Table 2: Oligonucleotide Primers for Amplification of the Putative Promoter of Mouse PEA3 38

LIST OF FIGURES

Figure

1. The 5' End of the Mouse PEA3 Gene 12

2. The pGL3 Reporter Vector 16

3. Example of PCR Products Separated by Agarose Gel Electrophoresis 40

4. Example of Restriction Endonuclease Analysis of the Putative PEA3 Promoter/Luciferase Constructs 43

5. Regions of the PEA3 Promoter Cloned into the pGL3 Reporter Vector
   a. 5' Deletions of the Putative Promoter 46
   b. 3' Deletions of the Putative Promoter 47

6. Luciferase Activity of PEA3 Promoter/Luciferase Constructs
   a. 5' Deletions of the Putative Promoter 50
   b. 3' Deletions of the Putative Promoter 51

7. Luciferase Activity of PEA3 Promoter/Luciferase/SV40 Enhancer Constructs 56

8. Human PEA3 cDNA Probes 59

9. Southern Blot Analyses of Human Genomic Cosmid Isolates
   a. Hybridization of a 3' Probe to hPEA3g.4 and hPEA3g.7 63
   b. Hybridization of a 5' Probe to Cosmids Overlapping the 5' End of PEA3 64
   c. Hybridization of a 3' Probe to Cosmids Overlapping the 5'End of PEA3 65

10. Human Cosmid Isolates - A Summary 67

11. Subclones of Human Genomic PEA3 70

12. FISH Chromosomal Localization Results
    a. Localization of Human PEA3 to 17q21.3 72
    b. Localization of Mouse PEA3 to 11d 72

13. Genomic Alignment of the Putative Promoter Regions of Mouse and Human PEA3 75

14. Consensus Transcription Factor Binding Sites Conserved Between the Putative Promoter Regions of Mouse and Human PEA3

a. Detailed Map of Conserved Consensus Sequences 82

b. Schematic Map of Conserved Consensus Sequences 85

LIST OF APPENDICES

1. Examples of Transcription Factor Consensus Matrices 114

Matrices were downloaded from the Transfac® database. The matrix contains the position in the consensus and a value is given for each nucleotide.

2. Genomic Human PEA3 Sequence 117

PEA3 sequence starting at three kilobasepairs upstream from the putative transcription start site is presented. Exon sequences are underlined.

ABREVIATIONS

AMPGD	(3-(4-methoxyspiro[1,2-dioxetane-3,2'-tricyclo[3.3.1.1^{3,7}]decan]-4-yl)phenyl -β-D-galactopyranoside)
bp	base pair(s)
cfu	colony forming unit
dATP	deoxyadenosine triphosphate
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
dTTP	deoxythymidine triphosphate
DTT	dithiothreitol
E. coli	*Escherichia Coli*
EDTA	ethylenediaminetetraacetic acid
FISH	fluorescence *in situ* hybridization
luc	luciferase
nt	nucleotide(s)
O.D.	optical density
PCR	polymerase chain reaction
RNA	ribonucleic acid
RNAP	DNA-dependent RNA polymerase
Taq	*Thermophile aquaticus*
TAE	Tris-Acetate-EDTA
TE	Tris-EDTA

INTRODUCTION

PEA3 is a member of the *Ets* oncogene family

PEA3 (polyomavirus enhancer activator 3) is a nuclear protein, originally isolated from mouse 3T6 cells that is capable of binding to the PEA3 motif ($^{5'}$AGGAAG$^{3'}$) in the polyomavirus enhancer (Martin *et al.*, 1988). A cDNA encoding a protein capable of binding to the PEA3 motif was isolated from a mouse FM3A mammary tumour cell line and named PEA3 (Xin *et al.*, 1992). The PEA3 cDNA encodes a protein comprising 480 amino acids that migrates in denaturing polyacrylamide gels with an apparent molecular mass of 66 kDa. Antibodies to PEA3 identify three isoforms corresponding in mass to 66, 60 and 58 kDa in all mammalian cell lines characterized to date.

Sequence analyses of the cDNA reveals that PEA3 is a member of the Ets proto-oncogene family. Ets proteins are transcriptional regulatory proteins and are involved in many biological processes including T-cell activation, growth control, development and transformation (Wasylyk *et al.*, 1993). All Ets genes possess highly-related ETS domains comprising ~85 amino acids that are sufficient for sequence-specific DNA binding. Ets proteins bind to 10 base pair sequence elements; a common feature of such Ets-binding sites is a central 5'-GGA,A/T-3' motif (Wasylyk *et al.*, 1993). Specificity for binding by particular Ets proteins is conferred by sequences flanking this central core. The ETS domain contains three essentially invariant tryptophans each separated by 17-21 amino acids, followed by a C-terminal rich in basic amino acids (Janhnecht and Nordheim, 1993).

The ETS domain for Ets-1 binds to the major groove of DNA as a winged helix-turn-helix motif. This motif consists of three α-helices packed onto a 4 strand anti-parallel β-sheet (Donaldson *et al.*, 1996).

The founding member of the Ets proto-oncogene family, v-*ets*, discovered as part of a fusion protein with *gag* and *myb*, is expressed by the avian E26 erythroblastosis virus (Leprince *et al.*, 1983). The term, ETS, originates from E26 transformation-specific, and is given to the domain of cellular proteins which resemble the v-Ets protein. In the past decade, over 30 ets-related proteins have been discovered in species ranging from human to *Drosophila*. In vertebrates, the *ets* gene family includes c-*ets*-1 (Watson *et al.*, 1988), c-*ets*-2 (Watson *et al.*, 1988), *ER71* (Brown and McKnight, 1992), *ERG* (Reddy *et al.*, 1987; Rao *et al.*, 1987), *Fli-1* (Ben-David *et al.*, 1991), *Sap*-1 (Dalton & Treisman, 1992), *ERP* (Lopez *et al.*, 1994), *Elk*-1 (Rao *et al.*, 1989), *GABPα* (LaMarco *et al.*, 1991), *FEV* (Peter *et al.*,1997), *TEL* (Golub *et al.*,1994), *Elf-1* (Thompson *et al.*, 1992), *PEA3/E1AF* (Xin *et al.*, 1992; Higashino *et al.*, 1993), *ER81/ETV1* (Brown and McKnight, 1992; Monté *et al.*, 1995), *ERM* (Monté *et al*, 1994), *NERF* (Oettgen *et al.*, 1996), *PU.1/Spi*-1 (Klemsz *et al.*, 1990; Goebl *et al.*, 1990), *SPI-B* (Ray *et al.*, 1992), *ERF* (Liu *et al.*, 1997) and *ESX* (Chang *et al.*, 1997).

Ets proteins can be grouped into subfamilies based upon position and sequence similarity of the ETS domain. PEA3 is the founding member of the PEA3 subfamily, which includes ER81 (Brown and McKnight, 1992; Monté *et al.*, 1995; Jeon *et al.*, 1995) and ERM (Monté *et al.*, 1994). The ETS domains of these three proteins are 95% identical and share additional sequences over their entire length. The tissue expression

patterns of each of the PEA3 subfamily members is summarized in Table 1. In the adult mouse, PEA3 RNA is most abundant in the brain and epididymis and to a lesser extent in testis, mammary gland and ovary (Xin *et al.*, 1992). PEA3 mRNA has recently been detected in the uterus (MacNeil and Hassell, personal communication). In normal human tissues, ERM mRNA is highly expressed in the brain, placenta, and lung and expressed to a lower extent in muscle, kidney, pancreas and heart. ERM RNA, in mouse, is detected in the spleen, testis, kidney and brain (Monté *et al.*, 1994). In human tissues, ER81 mRNA is expressed highly in testis, lung and heart, moderately in colon, pancreas, spleen and small intestine, weakly in thymus, liver and prostate and very weakly in skeletal muscle, ovary and kidney (Monté *et al.*, 1995).

Ets proteins and oncogenesis

Members of the Ets family are involved in oncogenesis both in animals and humans. In addition to v-*ets*1, which causes erythroleukemias in chickens (Leprince *et al.*, 1983), *Spi*-1/*PU.1* and *Fli*-1 transcripts are activated by retroviral insertions induced by the spleen focus-forming and Friend viruses, respectively (Ben-David *et al.*, 1991; Moreau-Gachelin *et al.*, 1988) and result in erythroleukemias. In humans, chromosomal translocations, which fuse a variety of ETS DNA-binding domains with the N-terminal domain of *EWS* give rise to Ewing's sarcomas. Almost 90% of Ewing's tumors are characterized by a t(11;22)(q24;q12) that fuses the *EWS* gene on chromosome 22 to *Fli*-1 on chromosome ll. This translocation produces a chimeric protein, which acts as a more potent transcriptional activator than FLI-1 (May *et al.*, 1993; Ohno *et al.*, 1993; Bailly *et*

Table 1. mRNA Expression of PEA3 Subfamily Members

The relative levels of mRNA expression of each gene in each tissue is shown. Arbitrary quantification of expression is as follows: none: -, very weak: ❖, weak: ❖❖, moderate: ❖❖❖ , high: ❖❖❖❖ and very high: ❖❖❖❖❖. Tissues which have not been analyzed are indicated by N/A.

	mPEA3	mERM	hERM	mER81	hER81
Brain	◆◆◆	◆◆◆◆◆	◆◆◆◆◆	◆◆◆◆◆	◆◆◆◆◆
Lung	-	N/A	◆◆◆◆	N/A	◆◆◆◆
Skeletal Muscle	-	N/A	◆◆◆	N/A	◆
Kidney	-	◆◆◆	◆◆	◆◆◆◆◆	◆
Liver	-	N/A	◆	N/A	◆◆
Pancreas	-	N/A	◆◆◆◆	N/A	◆◆◆
Spleen	-	◆◆◆	N/A	N/A	◆◆◆
Thymus	-	N/A	N/A	N/A	◆◆◆
Heart	-	N/A	◆◆◆	N/A	◆◆◆◆
Colon	N/A	N/A	◆◆◆	N/A	◆◆◆
Small Intestine	N/A	N/A	N/A	N/A	◆◆◆
Epididymis	◆◆	N/A	N/A	N/A	N/A
Prostate	N/A	N/A	N/A	N/A	◆◆
Testes	◆	◆◆◆	N/A	◆◆◆	◆◆◆◆
Mammary gland	◆	N/A	N/A	N/A	N/A
Uterus	◆◆	N/A	N/A	N/A	N/A
Ovary	◆◆	N/A	-	N/A	◆
Placenta	N/A	N/A	◆◆◆◆◆	N/A	-

al.,1994) suggesting that increased expression of its target genes is required for tumorigenesis (Lessnick *et al.*, 1995). Fusion of *ERG* on chromosome 21 to *EWS* sequences occurs in approximately 10% of Ewing's tumors (Zucman *et al.*, 1993; Sorensen *et al.*, 1994). In rare cases of Ewing's tumors, *EWS* can be fused to *PEA3* on chromosome 17 or *ETV1* on chromosome 7 (Kaneko *et al.*, 1996; Urano *et al.*, 1996; Jeon *et al.*, 1995). In some myeloid leukemias, the *ERG* gene is fused to *TLS* on chromosome 16 (Ichikawa *et al.*, 1994). The *TEL* gene on chromosome 12 undergoes a number of translocations giving rise to a variety of leukemias (Golub *et al.*, 1994, 1995; Buijs *et al.*, 1995; Romana *et al.*, 1995; Papadopoulos *et al.*, 1995).

Ets proteins can also play indirect roles in oncogenesis. Ets proteins are downstream targets of constitutively-activated non-nuclear oncoproteins. The *HER-2/neu* gene encodes a 185 kDa receptor tyrosine kinase that is structurally related to the epidermal growth factor receptor (EGFR) (Singleton and Strickler, 1992). Overexpression of this gene is involved in 20-30% of all breast carcinomas which have the inclination to invade and metastasize to other tissues leading to a poor prognosis for the patient (Slamon *et al.*, 1987). Amplification of the *HER-2/neu* gene on chromosome 17 accounts for only part of the overexpression. Human breast tumor-derived cell lines overexpress the mRNA 6-8 fold per gene copy compared to normal mammary cell lines regardless of gene amplification (King *et al.*, 1989; Inglehart *et al.*, 1990; Parkes *et al.*, 1990). Interestingly, the mRNA for PEA3 is overexpressed in mammary tumors of transgenic mice that overexpress the rat *neu* gene in their mammary epithelium (Trimble *et al.*, 1993). The primary and metastatic mammary tumors all overexpress PEA3 mRNA,

but the normal tissues adjacent to the tumors do not. At present, there is no evidence to suggest that the PEA3 gene is amplified in these tumours suggesting that PEA3 expression is controlled at the level of transcription and may be a specific downstream target of Neu. Furthermore, endogenous neu mRNA is also elevated in mammary tumors overexpressing PEA3 (Trimble *et al.*, 1993). Perhaps an event initially enhances the Neu tyrosine kinase activity, which elevates the expression of PEA3. Increased synthesis of PEA3 transcripts lead to elevated expression of both Neu mRNA and target genes of PEA3 (Benz *et al.*, *Oncogene*, in press).

There is evidence that PEA3 may be a transcriptional activator of those proteins which are involved in degradation of the extracellular matrix and thus impart a metastatic and invasive phenotype to the tumor cells. Consensus PEA3 binding sites occur in the promoters of matrix metalloproteinases (MMPs) such as collagenase (MMP-1), stromelysin-1 (MMP-3), and gelatinase B (MMP-9). In transient transfection assays, PEA3 transactivates these promoters (Higashino, 1995). PEA3 binding sites also occur in the promoters of urokinase type plasminogen activator (Rorth *et al.*, 1990; Nerlov *et al.*, 1992), stromelysin-2 (MMP-10), stromelysin-3 (MMP-11) and matrilysin (MMP-7) (Matrisian *et al.*, 1994). In addition, exogenous expression of PEA3 in the non-metastatic breast cancer cell line, MCF-7, increases the cell's invasive and metastatic properties (Kaya *et al.*, 1996). There are obviously compelling reasons to establish the role of PEA3 upregulation in human breast tumors. An increase in PEA3 gene transcription could lead to an increase in PEA3 protein. One possible outcome might be an increase in expression of downstream target genes, such as the matrix metalloproteinases, potentially leading to

degradation of the cell matrix and metastasis of tumor cells. HER-2 is also a downstream target of PEA3; increased expression of HER-2 may lead to transformation. Furthermore, there is evidence that PEA3 autoregulates its own promoter (Benz *et al., Oncogene,* in press).

Initiation of transcription by RNA Polymerase II

Transcription initiation is one of the most important ways in which gene expression is regulated. It is crucial for the cell to regulate the initiation of gene expression accurately and selectively. RNA polymerase II (RNAP II) transcribes genes which encode proteins and other small RNAs (snRNAs) with the exception of the U6 RNA. There are two possible core elements which can be used independently or together for directing the initiation of transcription. One, referred to as the TATA box, is usually centered 30 nucleotides upstream of the transcription start site. The other, known as an initiator (Inr) element, encompasses the transcription start site (Smale and Baltimore, 1989). For specificity of initiation, eukaryotic RNA polymerases require accessory proteins known as the general transcription factors (GTFs). The seven GTFs are TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH, and TFIIJ (Zawel and Reinberg, 1993). All promoters transcribed by RNAP II are recognized specifically by a DNA-binding transcription factor, TFIID. This factor comprises many subunits; a 38 kD polypeptide, which binds to the TATA element (TBP; TATA binding protein) and TBP-associated factors (TAFs) (Peterson et al., 1990; Hoffman et al., 1990).

For transcription on TATA-containing promoters, the first step is the formation of the preinitiation complex which is the binding of TFIID to the TATA element (Burtowski *et al.*, 1989; Maldonado *et al.*, 1990). Subsequently, this site of nucleation allows two other factors, TFIIA and TFIIB, to associate with TFIID at the promoter. Following this event, RNAP II is recruited to the promoter along with TFIIF. Binding of TFIIE, TFIIH and TFIIJ completes assembly of the preinitiation complex (Zawel and Reinberg, 1993). Following assembly of this complex, an event known as 'promoter clearance' occurs and elongation of the mRNA takes place.

For promoters lacking the TATA element, regulatory transcription factors apparently tether TFIID to the promoter. TATA-less promoters contain all the information necessary for determining specific initiation of transcription *in vivo* and *in vitro*. Various Inr elements have been described and classified according to sequence homology. These are the sequences between -6 and +11 in the major late promoter of adenovirus 2 (Smale and Balitmore, 1989), between -3 and +5 in the promoter of terminal deoxytransferase (TdT) (Smale and Baltimore, 1989; Smale *et al.*, 1990), between -2 and +10 for erythroid-specific promoter of the human porphobilinogen deaminase gene (Beaupain *et al.*, 1990), between -6 to +11 in the P5 promoter of adeno-associated virus type 2 (Seto *et al.*, 1991), and between -11 and +9 in promoter of the mouse dihydrofolate reductase gene (Means and Farnham, 1990). There is a close correlation between the transcriptional activity of these sequences and the similarity of their start site to the Inr consensus, $_{-2}$KCABHYBY$_{+6}$, where K = G or T, B = C, G or T, H = A, C or T, Y = C or T (Bucher, 1990). Inr binding

proteins specifically interact with the Inr and tether TFIID so that a transcription competent complex can be formed. TFIID may also bind weakly to the sequence around -30; this is stabilized by activator proteins bound to adjacent elements. The TAFs are necessary for the interaction between TFIID and activator proteins (Pugh and Tjian, 1992). As with the initiation process on the TATA containing promoters, TFIID assembles the remaining basal factors or GTFs. At a later step, the RNAP II interacts with the Inr. The Inr is critical for positioning of the polymerase.

The putative Inr element at the major transcription start site for PEA3 is ($_{-3}$CTCA CAACT$_{+6}$) which has only one mismatch to the Inr consensus sequence. The sequence encompassing the start site defining exon 1' ($_{-2}$CCAGTTGG$_{+6}$) has two mismatches to the consensus. Both of the these sequences are considered to be very significant since they both have a percentage match which is above 81% (Bucher, 1990).

The Luciferase Expression System

Several reporter genes are used to evaluate gene regulation and expression in mammalian cells, including chloramphenicol acetyltransferase (*CAT*), β-galactosidase (*lac Z*), secreted alkaline phosphatase (*SEAP*) and luciferase (*luc*) (Gorman, 1985; Alam and Cook, 1990; Henthorn et al., 1988; de Wet *et al.*, 1985). A recent trend has been to use chemiluminescent assays for reporter gene products because of the sensitivity and speed of these non-isotopic detection techniques. The luciferase gene, cloned from the North

American firefly *Photinus pyralis* (de Wet *et al.*, 1985) is an excellent choice for the reporting of transcriptional activity in eukaryotic cells. Luciferase generates luminescence through mono-oxygenation of luciferin, utilizing O_2 and ATP as co-substrates.

$$\text{luciferase} + \text{luciferin} + \text{ATP} \xleftrightarrow{Mg^{2+}} \text{luciferase}\bullet\text{luciferyl-AMP} + \text{PPi}$$

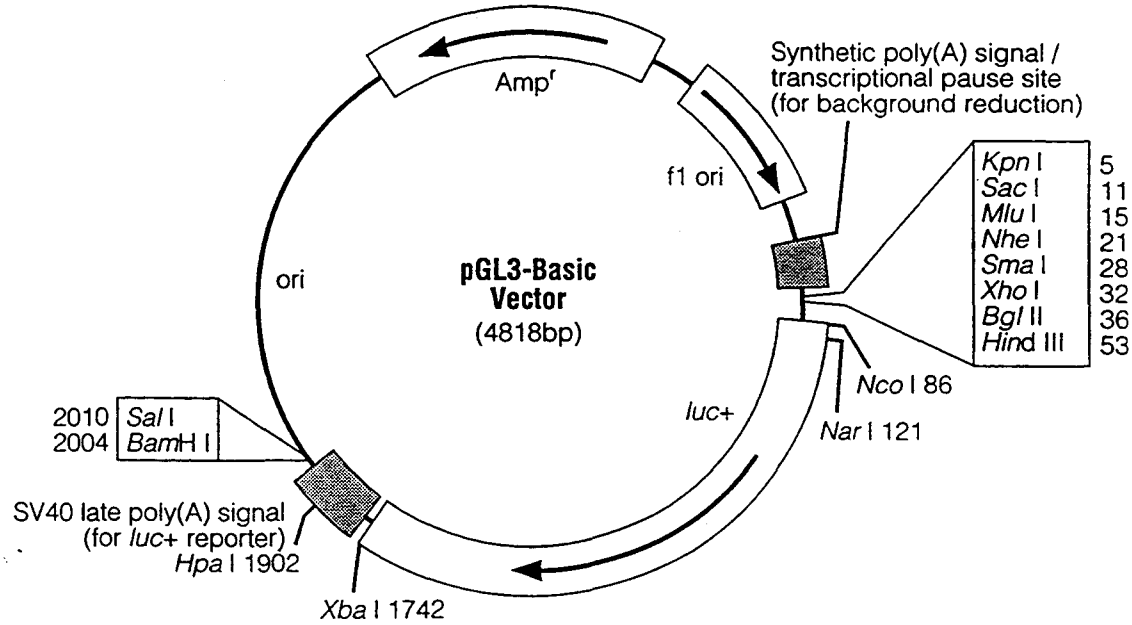
$$\text{luciferase}\bullet\text{luciferyl-AMP} + O_2 \rightarrow \text{luciferase} + \text{oxyluciferin} + \text{AMP} + CO_2 + h\nu$$

Light output is linearly proportional to the amount of luciferase. Therefore, luminescence correlates directly to expression of the reporter gene in transfected cells. The pGL3 vector series from Promega, provides an optimal system for investigating the molecular biology of gene expression. The pGL3 vector (Fig. 1) contains multiple cloning sites upstream of the luciferase reporter gene, and a downsteam polyadenylation signal for stable mRNA synthesis. A polyadenylation signal upstream from the cloning site functions as a transcriptional pause site to reduce background transcription. The modified sequence of the luciferase gene localizes it to the cytoplasm and the translational start site contains a Kozak consensus sequence to enhance translational efficiency (Kozak, 1991).

Transfection of a control reporter gene is often a control for transfection efficiency. β-galactosidase reporter plasmids are ideal for cotransfection with luciferase reporter constructs, since cell extracts can be prepared using the same lysis buffer and chemiluminescent assays can be performed (Jain *et al.*, 1991). The chemiluminescent method can detect 1 fg to 20 ng of purified luciferase or 10 fg to 20 ng of purified

Figure 1. The pGL3 Reporter Vector

The cloning site is illustrated in the box at the right. The luciferase gene is indicated by *luc*+ followed by the SV40 late poly(A) signal. The vector also possesses a synthetic poly(A) signal upstream of the luciferase gene, along with a transcriptional pause site, to reduce background luciferase expression. ori, origin of replication; Amp^r, β–lactamase gene for ampicillin resistance; f1 ori, f1 origin.


Amp^r
Synthetic poly(A) signal / transcriptional pause site (for background reduction)
f1 ori
Kpn I 5
Sac I 11
Mlu I 15
Nhe I 21
Sma I 28
Xho I 32
Bgl II 36
Hind III 53
ori
pGL3-Basic Vector
(4818bp)
Nco I 86
Nar I 121
luc+
2010 Sal I
2004 BamH I
SV40 late poly(A) signal (for luc+ reporter)
Hpa I 1902
Xba I 1742

β-galactosidase (Jain *et al.*, 1991). The substrate for the chemiluminescent β-galactosidase reaction is AMPGD (3- (4 - methoxyspiro [1,2 - dioxetane - 3,2' - tricyclo [3.3.1.1^{3,7}] decan] - 4 - yl) phenyl - β - D -galactopyranoside). This compound has a β-galactosidase moiety, that is cleaved by β-galactosidase to yield the dioxetane (AMPD) anion. The compound decomposes at a pH > 9 and yields adamantanone and the methyl metaoxybenzoate anion, which is in an excited state and emits light. Thus, light emission can be used to quantitate the hydrolysis of AMPGD by β-galactosidase (Jain *et al.*, 1991).

The Dual-Light™ reporter gene assay system, developed by Tropix, sequentially detects luciferase and β-galactosidase activity in a single extract sample. The assay uses the substrates luciferin and Galacton-plus™ for the detection of luciferase and β-galactosidase activities, respectively. Both substrates are injected into the sample at the same time and the light signals from the luciferase catalyzed reaction is measured immediately. The light signal produced by the luciferase reaction decays with a half life of approximately one minute. The products from the β-galactosidase reaction accumulate for 60 minutes at which time the Accelerator™ buffer is injected which raises the pH and increases the light intensity of the products. This light production is also measured with a luminometer. The β-galactosidase emission has a half life of up to 180 min.

Characterization of the Mouse PEA3 Gene Promoter

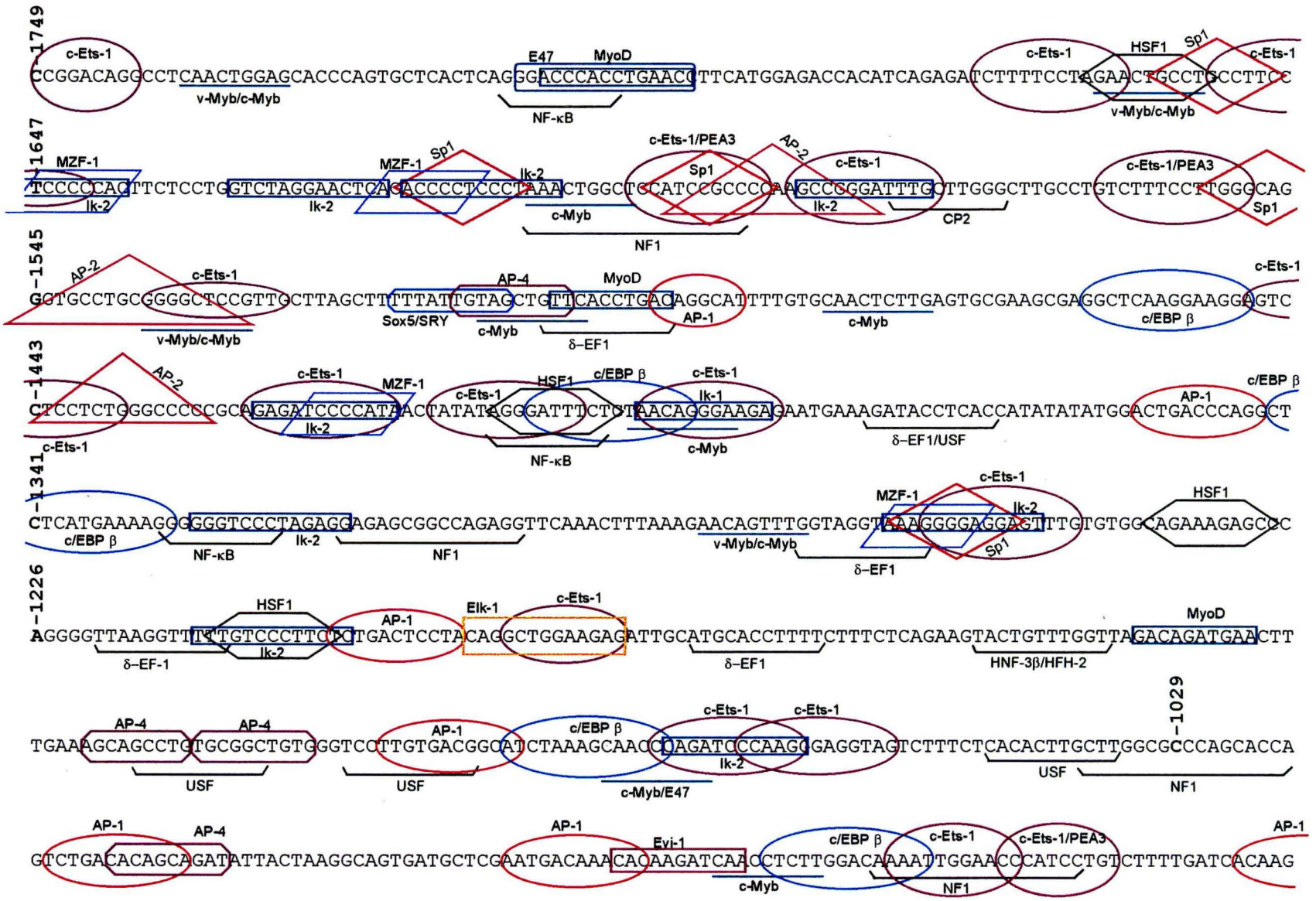
The main objective of this project was to map the sequences governing the expression of PEA3. To this end, a series of luciferase reporter plasmids containing

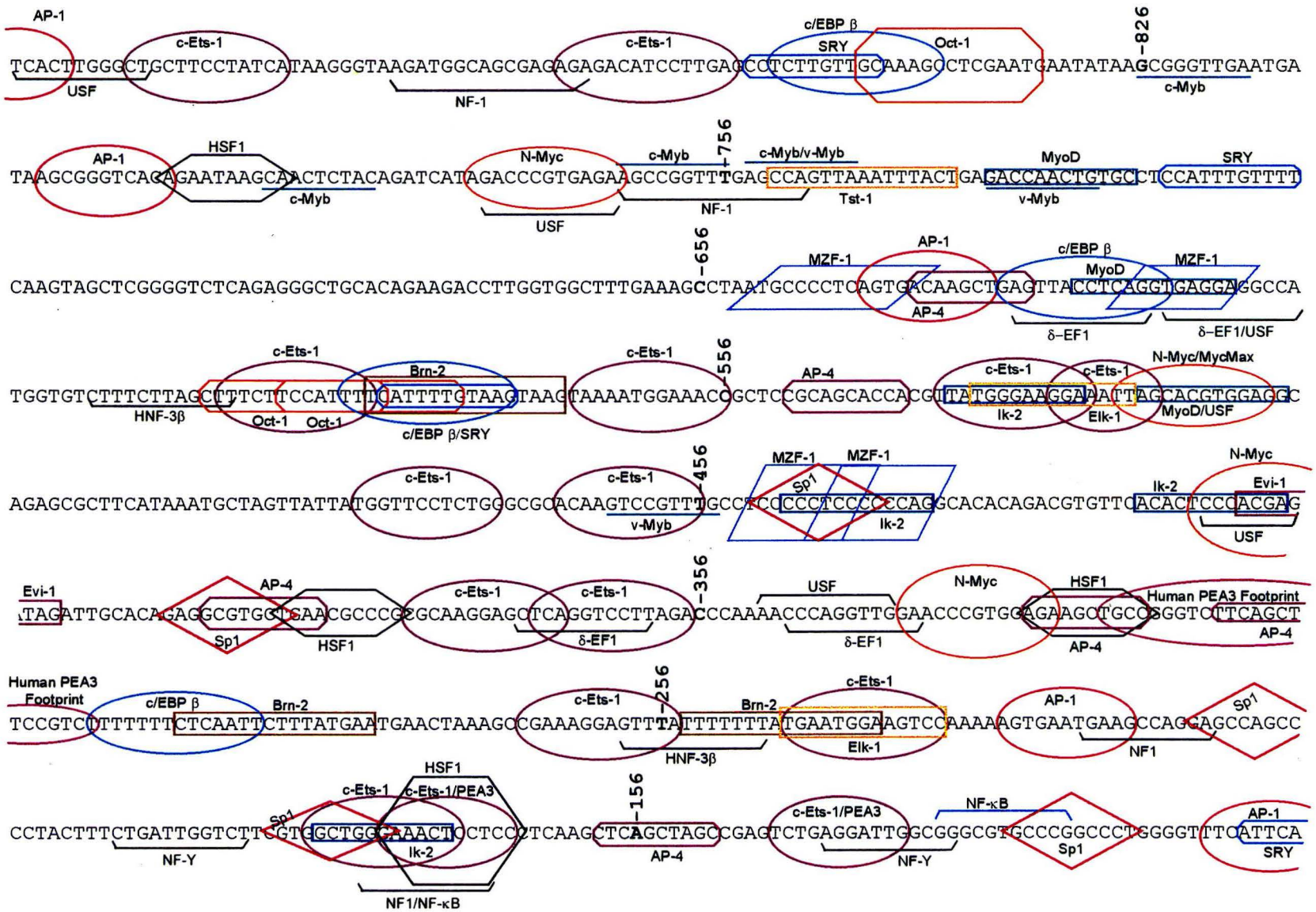
deletions of the putative PEA3 promoter were constructed. These reporter plasmids were then transiently transfected into the FM3A cell line which normally expresses high levels of PEA3 mRNA and protein (Xin *et al.,* 1992). The relative *in vivo* activities of these promoters was then evaluated.

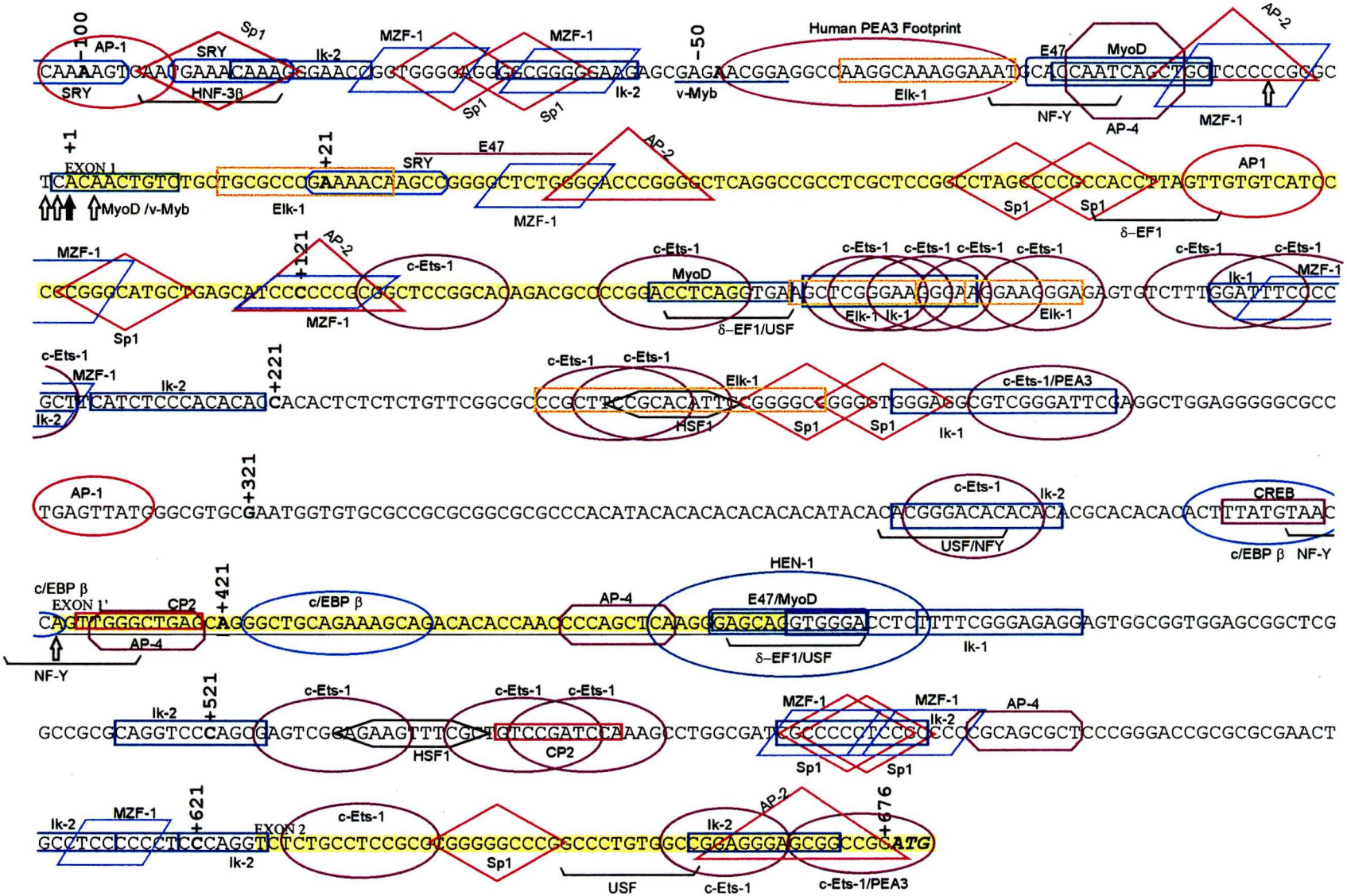
The PEA3 gene consists of 13 exons and spans 14 kb (Smillie, 1993). The sequence of the entire gene and 3 kb upstream of the first exon is known. There are multiple initiation sites located at the beginning of the first exon, as well as an alternate transcription start site at exon 1' (Laing *et al.,* personal communication). The major transcription initiation site is indicated by a solid arrow and the other initiation sites as hollow arrows (Fig. 2). The notion of the alternate promoters has been reported for numerous genes including the chicken c-ets-1 locus, which gives rise to two distinct mRNAs (Crepieux et al., 1993). The putative translation initiation codon for PEA3 gene occurs in exon 2 and appears in bold-face type (Fig. 2). The upstream region of the mouse PEA3 gene contains many putative transcription factor binding sites (Fig. 2), however, these sites are only theoretical. The sites were identified based on an 85% match to the consensus binding site from the TRANSFAC data base (Quandt *et al.,* 1995, Wingender *et al.,* 1996; Wingender *et al.,* 1997). Neither of the two putative PEA3 promoter regions has an identifiable TATA element but, as mentioned previously, both have consensus Inr elements surrounding the transcription initiation site. Additional evidence suggesting that the upstream region contains a functional promoter was based on another sequence analysis program called Gene Finder (http://defrag.bcm.tmc.edu:9503/gene-finder/gf.html). This program was used for the recognition of Pol II promoter regions and start sites of

Figure 2. The 5' End of the Mouse PEA3 Gene

The 5' end of the mouse PEA3 gene was analyzed for the presence of transcription factor consensus sequences using MatInspector® software. Putative recognition sequences matching 85% of the consensus are labelled. The sequence is numbered relative to the major transcription start site, +1, indicated by a solid arrow. Minor transcription start sites are indicated by hollow arrows. Exon sequences are underlined. The putative translation start codon is in boldface type.


-1749
c-Ets-1
CCGGACAGGCCTCAACTGGAGCACCCAGTGCTCACTCAGGGACCCACCTGAACCTTCATGGAGACCACATCAGAGATCTTTTCCTAGAACTGCCTGCCTTCC
v-Myb/c-Myb
E47
MyoD
NF-κB
c-Ets-1
HSF1
Sp1
c-Ets-1
v-Myb/c-Myb
-1647
MZF-1
TCCCCCACTTCTCCTGGTCTAGGAACTCACACCCCTCCCTAAACTGGCTGCATCCGCCCCAAGCCCGGATTTGCTTGGGCTTGCCTGTCTTTCCTTGGGCAG
Ik-2
Ik-2
MZF-1
Sp1
Ik-2
c-Myb
c-Ets-1/PEA3
Sp1
AP-2
c-Ets-1
Ik-2
CP2
NF1
c-Ets-1/PEA3
Sp1
-1545
AP-2
c-Ets-1
GGTGCCTGCGGGGCTCCGTTGCTTAGCTTTTTATTGTAGCTGTTCACCTGACAGGCATTTTGTGCAACTCTTGAGTGCGAAGCGAGGCTCAAGGAAGGAGTC
v-Myb/c-Myb
Sox5/SRY
AP-4
c-Myb
MyoD
AP-1
δ-EF1
c-Myb
c-Ets-1
c/EBP β
-1443
AP-2
CTCCTCTGGGCCCCCGCAGAGATCCCCATAACTATATAGGGATTTCTGTAACAGGGAAGAGAATGAAAGATACCTCACCATATATATGGACTGACCCAGGCT
c-Ets-1
c-Ets-1
MZF-1
Ik-2
c-Ets-1
HSF1
c/EBP β
c-Ets-1
Ik-1
NF-κB
c-Myb
δ-EF1/USF
AP-1
c/EBP β
-1341
CTCATGAAAAGGGGGGGTCCCTAGAGGAGAGCGGCCAGAGGTTCAAACTTTAAAGAACAGTTTGGTAGGTAAAGGGGAGGAGTTTGTGTGGCAGAAAGAGCCC
c/EBP β
NF-κB
Ik-2
NF1
v-Myb/c-Myb
MZF-1
c-Ets-1
Ik-2
Sp1
δ-EF1
HSF1
-1226
AGGGGTTAAGGTTTTTGTCCCTTCTCTGACTCCTACAGGCTGGAAGAGATTGCATGCACCTTTTCTTTCTCAGAAGTACTGTTTGGTTAGACAGATGAACTT
δ-EF-1
Ik-2
HSF1
AP-1
Elk-1
c-Ets-1
δ-EF1
HNF-3β/HFH-2
MyoD
-1029
AP-4
AP-4
TGAAAGCAGCCTGTGCGGCTGTGGGTCCTTGTGACGGCATCTAAAGCAACCCAGATCCCAAGGGAGGTAGTCTTTCTCACACTTGCTTGGCGCCCAGCACCA
USF
USF
AP-1
c/EBP β
c-Ets-1
c-Ets-1
Ik-2
c-Myb/E47
USF
NF1
AP-1
AP-4
GTCTGACACAGCAGATATTACTAAGGCAGTGATGCTCGAATGACAAACACAAGATCAACCTCTTGGACAAAATTGGAACCCATCCTGTCTTTTGATCACAAG
AP-1
Evi-1
c-Myb
c/EBP β
c-Ets-1
c-Ets-1/PEA3
NF1
AP-1


AP-1
c-Ets-1
USF
NF-1
c-Ets-1
c/EBP β
SRY
Oct-1
-826
c-Myb
AP-1
HSF1
c-Myb
N-Myc
USF
c-Myb
-756
c-Myb/v-Myb
NF-1
Tst-1
MyoD
v-Myb
SRY
-656
MZF-1
AP-1
AP-4
c/EBP β
MyoD
MZF-1
δ-EF1
δ-EF1/USF
c-Ets-1
HNF-3β
Oct-1
Oct-1
Brn-2
c/EBP β/SRY
c-Ets-1
-556
AP-4
c-Ets-1
Ik-2
c-Ets-1
Elk-1
N-Myc/MycMax
MyoD/USF
c-Ets-1
c-Ets-1
-456
v-Myb
MZF-1
MZF-1
Sp1
Ik-2
N-Myc
Ik-2
Evi-1
USF
Evi-1
AP-4
Sp1
HSF1
c-Ets-1
c-Ets-1
δ-EF1
-356
USF
δ-EF1
N-Myc
HSF1
AP-4
Human PEA3 Footprint
AP-4
Human PEA3 Footprint
c/EBP β
Brn-2
c-Ets-1
-256
Brn-2
HNF-3β
c-Ets-1
Elk-1
AP-1
NF1
Sp1
HSF1
c-Ets-1
c-Ets-1/PEA3
Sp1
Ik-2
NF-Y
NF1/NF-κB
-156
AP-4
c-Ets-1/PEA3
NF-Y
NF-κB
Sp1
AP-1
SRY


-100
AP-1
SRY
Sp1
Ik-2
MZF-1
MZF-1
-50
Human PEA3 Footprint
E47
MyoD
AP-2
SRY
HNF-3β
Sp1
Sp1
Ik-2
v-Myb
Elk-1
NF-Y
AP-4
MZF-1
+1
EXON 1
+21
SRY
E47
AP-2
AP1
MyoD /v-Myb
Elk-1
MZF-1
Sp1
Sp1
δ-EF1
MZF-1
+121
AP-2
c-Ets-1
c-Ets-1
MyoD
c-Ets-1
c-Ets-1
c-Ets-1
c-Ets-1
c-Ets-1
c-Ets-1
Ik-1
MZF-1
Sp1
MZF-1
δ-EF1/USF
Elk-1
Ik-1
Elk-1
c-Ets-1
MZF-1
+221
Ik-2
Ik-2
c-Ets-1
c-Ets-1
Elk-1
c-Ets-1/PEA3
HSF1
Sp1
Sp1
Ik-1
AP-1
+321
c-Ets-1
Ik-2
CREB
USF/NFY
c/EBP β
NF-Y
c/EBP β
EXON 1'
CP2
+421
c/EBP β
AP-4
HEN-1
E47/MyoD
AP-4
δ-EF1/USF
Ik-1
NF-Y
Ik-2
+521
c-Ets-1
c-Ets-1
c-Ets-1
HSF1
CP2
MZF-1
MZF-1
Ik-2
AP-4
Sp1
Sp1
Ik-2
MZF-1
+621
EXON 2
c-Ets-1
Ik-2
AP-2
+676
Ik-2
Sp1
USF
c-Ets-1
c-Ets-1/PEA3
ATG

transcription (Prestridge *et al.*, 1995). This algorithm predicts potential transcription start site positions based on the density of functional motifs, using a transcription factor database, as well as oligonucleotide composition at the predicted start site. The entire mouse genomic PEA3 sequence, containing the entire locus and 3 kb upstream from the putative start site, was analyzed using this program. Although this program did not predict the start sites which have been identified for PEA3 (Fig. 2), 4 sites were identified within 500 bp range and were centered around the first intron. The nearest predicted binding site was located 53 base pairs upstream from the apparent start site at exon 1. There was also a predicted start site located 58 base pairs downstream from the start site occurring at exon 1'. The high density of consensus binding sites in this region, relative to the rest of the genomic sequence, encouraged us to begin mapping functional promoter elements near this region.

The upstream region of the PEA3 gene has characteristics of a CpG island (Gardiner-Garden and Frommer, 1987). When the entire 21 kb sequence of genomic mouse PEA3 was searched for CpG islands using the Grail program (http://avalon.epm.ornl.gov/Grail-bin/GrailForm-post), a single CpG island was found to be in the region from 139 bp upstream from the major transcription initiation site to 1231 bp downstream. This CpG island contains a %GC of 64.6 and has a CpG score of 0.70.

In a parallel set of experiments , the promoter region for the human PEA3 gene was to be isolated and sequenced. This was performed to aid in the search for critical elements governing PEA3 expression. Furthermore, the chromosomal location of PEA3 was to be

mapped. By knowing the position of the PEA3 gene in the human genome this may facilitate the search for human malignancies with chromosome changes involving the PEA3 gene.

OBJECTIVES

1. To amplify 26 defined regions of the mouse PEA3 genome with flanking sequences on either side of the major transcription start site.
2. To isolate 9 restriction fragments of the mouse PEA3 genome with sequences flanking the major transcription start site.
3. To clone these putative promoter sequences into the pGL3-basic luciferase reporter vector in the correct orientation.
4. To assess their relative *in vivo* promoter activity using the luciferase assay following transient transfection into the FM3A cell line.
5. To determine the effect of an SV40 enhancer sequence placed 3 kb upstream of the PEA3 promoter in the luciferase reporter plasmids.
6. To isolate the human PEA3 gene from a genomic library and to sequence the putative promoter region.
7. To align the human and mouse genomic regions encompassing the putative promoter and assess these sequences for the presence of consensus transcription factor binding sites.

MATERIALS

Restriction endonucleases were purchased from Life Technologies, Burlington, Ontario, or Boehringer Mannheim, Laval, Quebec, and were used according to manufacturer's specifications. Ultrapure low melting temperature agarose, large fragment of DNA polymerase I, T4 DNA ligase, *Taq* DNA polymerase, 1 kb ladder and 100 bp ladder DNA markers and lipofectAMINE™ transfection reagent were also obtained from Life Technologies.

The pGL3 series of luciferase reporter vectors were purchased from Promega Corporation, Madison, Wisconsin, USA. The Dual-Light™ assay system was purchased from Tropix, Bedford, Massachusetts, USA.

The human fetal brain genomic library and the pBluescript KS II vector were obtained from Stratagene, La Jolla, California, USA. [α-^{32}P] dCTP and Hybond N nylon membranes were obtained from Amersham Corporation, Oakville, Ontario.

METHODS

1. Construction of Mouse PEA3/pGL3 Chimera

a. PCR Amplification of the Putative Promoter Region of Mouse PEA3

The polymerase chain reaction (PCR) allows the selection and subsequent amplification and cloning of DNA fragments which otherwise would not easily be isolated from restriction endonuclease digestion. The selection of oligonucleotide primers defines a segment of DNA to be amplified and selection was optimized using PCgene® DNA analysis software. Primers chosen from defined regions of the sequence were analyzed for factors which may inhibit amplification such as their melting temperature (Tm), their potential to dimerize or whether they possessed a secondary structure.

For amplification, PCR reactions contained 5U of *Taq* DNA polymerase, 100 ng of template DNA, 1 mM dNTPs, *Taq* polymerase buffer (50 mM KCl, 10 mM Tris pH 8.8, 1.0 or 1.5 mM $MgCl_2$, 0.01% (w/v) gelatin, 0.1% Triton X-100) and 100 pmol of each primer. Each reaction had a volume of 100 μl and amplification was performed using 25 cycles of the following protocol: denaturation of DNA strands at 94°C for 35 seconds, annealing of primers at 60°C for 35 seconds and DNA synthesis at 72°C for 35 seconds. For each set of reactions, a negative control amplification containing no DNA template was performed.

To confirm the size of the amplified product, 0.1 volumes of the reaction was electrophoresed at 90 V for two hours in a 1% agarose gel and compared to a DNA ladder marker.

b. Restriction Endonuclease Isolation of the Putative Promoter Region of Mouse PEA3

A 6.0 kb *Nhe* I fragment, which has a 3' endpoint 156 bp upstream from the putative transcription initiation start, was identified using Southern analysis with a probe specific to the region -1341 to -156. This 6.0 kb fragment was ligated in the correct orientation upstream of a PCR-generated PEA3 promoter construct, which contained the sequence from -156 to +676.

The presence of an *Nhe*I site at position -156 in the mouse genomic sequence was used to generate numerous putative promoter clones. By simply excising sequences using the *Nhe*I site upstream of the cloned promoter sequence and the *Nhe*I restriction site at position -156, sequence between these 2 sites were removed and the vector was religated. The resulting vectors will now contain PEA3 promoter sequence which have 5' endpoints 156 bp upstream from the start site of transcription.

c. Purification and Modification of PCR Amplified DNA

In order to ensure more efficient cloning, PCR oligonucleotide primers were designed so that the amplified fragments would contain sequences corresponding to restriction endonuclease cleavage sites. By selecting different restriction sites for each end of the product, the PCR fragment can be easily cloned in the desired orientation.

The amplified putative PEA3 promoter sequences were engineered to contain a *Bgl*II restriction site at the 5' end and a *Hin*dIII site at the 3' end. To prepare for digestion, the PCR reaction was extracted with an equal volume of phenol:chloroform (1:1). The DNA was precipitated with the addition of 0.1 volumes of 3 M sodium acetate

(pH 5.5) and two volumes of ice cold absolute ethanol. The DNA was pelleted in a microcentrifuge and the pellet was washed with 70% ethanol and resuspended in 30 μl of TE buffer (10 mM Tris (pH 8.0), 1 mM EDTA (pH 8.0)). Five μg of DNA was digested for 20 hours with 15 U of *Hin*dIII endonuclease according to manufacturer's specifications. For the subsequent digestion with *Bgl*II endonuclease (15 U), the NaCl concentration was raised from 50 mM to 100 mM by the addition of 0.1 volumes of 0.5 M NaCl. This reaction was allowed to proceed for 20 hours.

Following digestion, the DNA fragments were separated by electrophoresis at 60 V for three hours on 1% low melting point agarose. The band of the correct size was excised from the gel and its concentration was approximated by comparison to the intensity of ethidium bromide stained DNA markers.

d. Preparation of pGL3 vector DNA

A cohesive-end directional cloning strategy was used clone the PCR amplified promoter sequences into the pGL3 luciferase vectors. Two cloning sites in the vector were chosen to avoid the intervening polylinker upstream of the luciferase gene.

The luciferase reporter plasmids, pGL3-basic (2 μg) and the SV40 enhancer containing plasmid, pGL3-enhancer (2 μg) were linearized by digestion with 10 U of *Hin*dIII endonuclease following manufacturer's specifications. This linear fragment was then digestioned with *Bgl*II endonuclease (10 U) for 1 hour at 37°C after increasing the NaCl concentration of the reaction from 50 mM to 100 mM. The linearized vector was

purified by electrophoresis at 60 V for three hours on 1% low melting point agarose and the appropriate fragment excised from the gel.

e. Ligation Reaction

Putative promoter fragments were ligated into the pGL3 reporter vectors using the following modification of an "in-gel" cloning protocol (Struhl, 1985). The approximate molar ratio of promoter region to vector DNA was 10:1. Gel slices containing the DNAs to be cloned were incubated for ten minutes at 65°C. A gel volume corresponding to 500 ng of PCR amplified DNA was mixed with a gel fragment of vector DNA corresponding to 50 ng. The mixture was allowed to cool to 37°C at which time 8μL of 5X ligation buffer (0.66 M Tris pH 7.5, 50 mM $MgCl_2$, 50 mM DTT, 10 mM ATP) and 2 U of T4 DNA ligase was added. The reaction volume was increased to 40μl with de-ionized distilled H_2O and incubated for 16 hours at 22°C. Following the ligation, the reaction was heated to 65°C for 10 minutes and then diluted two-fold in TCM buffer (20 mM Tris pH 7.5, 20 mM $CaCl_2$, and 20 mM $MgCl_2$).

f. Transformation

The transformation method was essentially that described in the Promega Protocols and Applications guide, 1991. To prepare competent cells, 250 ml of LB (Luria-Bertani) medium (10 g/l bacto-tryptone, 5g/l yeast extract, 10 g/l NaCl) was inoculated with 1 ml of an overnight culture of *E. Coli* JM109 (*rec*A1 *end*A1 *syr*A96 *thi*-1 *hsd*R17 (r_K-m_K^+) *supE44 relA1* Δ(*lac-pro*AB), [F',traΔ36,proAB,*lac*I^qZΔM15]) cells. This culture was

grown until the $O.D._{600}$ reached 0.5 and was then chilled on ice for two hours. The cells were harvested by centrifugation at 2500 x g for 10 minutes and resuspended in 250 ml of ice cold trituration buffer (100 mM $CaCl_2$, 70 mM $MgCl_2$, 40 mM sodium acetate (pH 5.5)). The cells were left on ice for 45 minutes, collected by centrifugation at 1800 x g for 10 minutes, resuspended in 25 ml of trituration buffer and 80% glycerol was added dropwise to a final concentration of 15% (v/v). Aliquots of the competent cells were frozen at -80°C.

For transformation, 200 µl of competent cells were thawed on ice and 10 µl of the ligation reaction was added. After incubation on ice for 30 minutes, the mixture was heat shocked at 42°C for 45 seconds followed by a 2 minute incubation on ice. Four volumes of S.O.C. medium (2% bacto-tryptone, 0.5% bacto-yeast extract, 10 mM NaCl, 2.5 mM KCl, 20 mM glucose, 20 mM $MgCl_2$) was added and the transformed cells incubated with gentle shaking for one hour at 37°C. Cells were plated on LB plates containing 50 µg/ml ampicillin and incubated for 16 hours at 37°C.

2. Identification and Characterization of Positive Chimera

a. Restriction Endonuclease Analysis of Positive Clones

Colonies which grew on the ampicillin plates were transferred with a sterile pipette tip to separate tubes containing 5 ml of LB medium containing 50 µg/ml of ampicillin. After a 16 hour incubation at 37°C with gentle shaking, a small scale plasmid isolation was performed to isolate plasmid DNA (Sambrook *et al.*, 1989). The purified DNAs were digested with *Hin*dIII and *Bgl*II endonucleases according to manufacturer's specifications

and the resulting fragments were separated by electrophoresis on a 1% agarose gel at 90 V together with a DNA ladder marker to confirm their sizes.

b. DNA Sequencing with GL2 and RV3 Sequencing Primers

The DNA sequence and orientation of the inserts in the chimera were determined by automated DNA sequence analysis using primers upstream and downstream of the cloning site. Sequencing reactions were performed in the Central Facility at the Institute for Molecular Biology and Biotechnology and analyzed with an ABI 373 automated DNA sequencing apparatus.

3. Luciferase Reporter Assay

a. Large Scale Preparation of Luciferase Constructs

For all transient transfections, the DNA was isolated by the alkaline lysis protocol as described (Sambrook *et al.,* 1989), and purified further by a CsCl density gradient centrifugation.

b. Transfection of FM3A Cells

The FM3A mouse mammary carcinoma cell line was grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS), gentamycin (5 μg/ml) and fungizone® (amphotericin B, 0.5 μg/ml). Cells were cultured at 37°C in humidified atmosphere containing 5% CO_2.

Cells were plated at a density of 5.0×10^5 cells/35 mm petri dish in antibiotic-free DMEM supplemented with 10% FBS 18 hours prior to transfection. For each transfection,

0.8 μg of experimental reporter DNA and 0.1 μg of the internal control plasmid, pSV-βgal, was mixed with 12 μg (6μl) of lipofectAMINE™ reagent in 200 μl of serum-free DMEM and incubated for 15 minutes at room temperature to allow DNA-lipid complexes to form. The volume was then increased to 1 ml with the addition of serum-free DMEM. Cells were washed twice with antibiotic and serum-free DMEM, overlaid with the DNA-lipid complexes and allowed to incubate at 37°C for 5 hours. At this time, 1 ml of DMEM supplemented with 20% FBS was added to each 35 mm petri dish. The cells were incubated for 24 hours and then cell extracts were prepared.

c. Preparation of Cell Extracts

To isolate cell extracts, cells growing on petri dishes were washed twice with cold 1X phosphate buffered saline (PBS) followed by the addition of 150 μl of cold, fresh lysis solution (100 mM potassium phosphate pH 7.8, 0.2% Triton X-100, 0.5 mM DTT). The cell lysate was removed from the plate with a cell scraper and transferred to a microcentrifuge tube and centrifuged to pellet nuclei. The supernatant was transferred to a fresh microcentrifuge tube and either assayed immediately or stored at -80°C for later use.

d. Luciferase and β-Galactosidase Assays

The Dual-Light™ assay system enables luciferase and β-galactosidase activity to be measured in the same cell extract. To perform the assay, 10 μl of cell extract was placed into luminometer sample tubes and mixed with 25 μl of a buffer containing the luciferase reaction reagents, minus the substrate, luciferin. The Berthold lumat 9501 luminometer,

which has an automatic injector, was used to inject 100 μl of a buffer containing luciferin and the β-galactosidase substrate, Galacton-Plus®, into the sample tube. The luciferase light signal was measured for five seconds after a two second delay. Each of the samples were incubated at room temperature for exactly 60 minutes after the addition of the Galacton-Plus® substrate before the automatic injection of 100 μl of Accelerator-II®. Following a two second delay, the light signal from the β-galactosidase reaction was measured for five seconds.

Luciferase activity for each of the putative promoter clones was calculated using the following formula:

$$\text{Luciferase Activity} = \frac{\text{(Luciferase Light Units - Background Light Units)}}{\text{(}\beta\text{-gal Light Units - Background }\beta\text{-gal Light Units)}} \Big/ \text{mol. DNA}$$

Luciferase activity was expressed relative to the control luciferase plasmid, pGL3-promoter, set to 100 luciferase units. This luciferase reporter plasmid is driven by the SV40 early promoter.

4. Statistical Analysis

In order to assess whether there was a statistically significant difference between luciferase activity of the chimera, a single-factor analysis of variance (ANOVA) test was performed (Montgomery, 1991). This test determines whether sets of data could have been drawn from the same sample. The data obtained from the luciferase assays was analyzed for variance using the ANOVA subprogram of Microsoft Excel®. This algorithm calculated the value of the test statistic for each pair-wise comparison of data which was

then compared to a critical value, such that, at a confidence level of 95%, it could be stated that the means of the two sets of data were, in fact, different.

5. Isolation and Characterization of Human PEA3 Cosmids

a. Human Cosmid Library Screening

In order to obtain the putative promoter sequence of the human PEA3 gene, a human fetal brain genomic cosmid library was screened. The library was plated onto 10 Hybond N nylon membranes overlaid upon 150 mm LB/kanamycin (50 μg/ml) plates at a concentration of 1.0×10^5 cfu/plate and the plates were incubated for 12 hours at 37°C. Replica membranes were made in duplicate by firmly pressing a fresh Hybond N membrane to the colonies on the master plates. The replicas were placed on fresh LB/kanamycin plates and were allowed to grow for 8 hours at 37°C. At this point, the cosmid DNA was isolated and prepared for hybridization. Each nylon membrane was placed for 30 seconds on top of a piece of Whatman 3MM paper soaked in 0.5 N NaOH. Following this, each membrane was placed on a solution of 1 M Tris (pH 7.5) for 30 seconds and finally on a solution of 0.5 M Tris (pH 7.5), 1.5 M NaCl for 30 seconds. The denatured DNA was then cross-linked to the membrane using 1200 μJ of UV radiation in a Stratagene UV Stratalinker®. Bacterial debris was removed with gentle washing in a solution of 0.5 M Tris (pH 7.5),1.5 M NaCl. The membranes were then placed in pre-hybridization buffer (6X SSC, 20 mM NaH_2PO_4, 0.4% SDS, 5X Denhardt's, denatured sheared herring sperm DNA (100 μg/ml)) for 2 hours at 65°C. At this point, fresh hybridization solution containing the probe at 5.0×10^5 cpm/ml was exchanged for the pre-hybridization buffer.

The probe hybridized to the membrane with gentle shaking in a water bath at 65°C for 16 hours. The membranes were washed three times for 15 minutes each at room temperature in a 2X SSC, 0.5% SDS solution. The membranes were air dried and exposed to Kodak XAR film with intensifying screens at -80°C for 72 hours. Once the films were developed, overlapping positive signals on the duplicate membranes were used to select colonies to be characterized further. At this stage a secondary screen was performed which involved repeating the above protocol until single colonies were chosen for Southern analyses.

Probes for cosmid screening hybridizations were prepared by the random primer technique (Feinberg and Volgelstein, 1983). DNA to be used as a probe was isolated from the vector sequences using restriction endonucleases and the DNA fragments were separated by electrophoresis at 90 V for 1 hour on 1% agarose. The probe DNA (500 ng) was extracted using a gel extraction column (Qiagen) and resuspended to a concentration of 50 ng/μl. The [α-^{32}P]-dCTP radiolabelled probe was prepared by mixing 50 ng probe DNA and 100 ng of random hexamers (Pharmacia) in a volume of 5 μl and then boiling for 10 minutes. The denatured DNA fragments were then placed on ice for 5 minutes to allow the hexamers to anneal to the single-stranded probe. To incorporate [α-^{32}P] dCTP, the reaction consisted of 90 mM HEPES (pH 6.6), 10 mM $MgCl_2$, 2 mM DTT, 1 mM dATP, 1 mM dTTP, 1 mM dGTP, 30 μCi of [α-^{32}P] dCTP and 6 U of large fragment DNA polymerase I. After a two hour incubation at 37°C the probe was purified using a Saphadex G-50 column and the incorporation of ^{32}P-dCTP was evaluated before hybridization.

b. Restriction Endonuclease and Southern Blot Analysis of Human PEA3 Cosmids

In order to characterize cosmids isolated from the human genomic library, restriction endonuclease analysis and Southern analysis (Southern, 1975) was performed to identify fragments of DNA sharing sequences with the probe. Agarose gels (0.7%) were used to separate DNA fragments after restriction endonuclease digestion. The fragments in the gel were denatured by soaking the gel in several volumes of 1.5 M NaCl, 0.5 N NaOH and then neutralized in a solution of 1 M Tris (pH 7.5), 1.5 M NaCl. The DNA was then transferred to Hybond N nylon membranes by capillary action (Sambrook *et al.*, 1989) and then crosslinked using 1200 μJ of UV radiation. Hybridization and washing were performed as outlined previously in the library screening protocol (section 5 a.). The membranes were exposed to Kodak XAR film with intensifying screens for 1 hour at -80°C.

c. Subcloning of DNA Fragments from the Human PEA3 Locus

In order to obtain constructs containing smaller regions of human genomic PEA3, fragments of the cosmid which shared sequences with the human PEA3 cDNA probes were subcloned into the pBluescript KS II vector. For these cloning experiments, the vector DNA was linearized using a single restriction endonuclease. To prevent religation of the plasmid, the DNA was dephosphorylated using 5 U of calf intestinal alkaline phosphatase in phosphatase buffer (50 mM Tris pH 7.5, 5 mM $MgCl_2$, 1 mM $ZnCl_2$) for 30 minutes at 37°C. The ligation reaction was performed as previously described in section 1 e.

d. Chromosomal Mapping of PEA3 to Human Chromosomes by FISH

The Chromosomal location of Human PEA3 was determined by Fluorescence In Situ Hybridization (FISH) (Lichter *et al.*, 1990) which was performed by the Human Genome FISH Mapping Resource Centre, Toronto.

Human lymphocyte chromosomes were stained with 4',6-diamindin-2-phenylindol-dihdrochoride (DAPI) (Heng and Tsui, 1993) and then hybridized to genomic fragments of human PEA3 labelled with biotin-11-dUTP. Avidin-fluorescein isothiocyanate (FITC) was used to detect the biotinylated probe and separate images of the DAPI banded chomosomes (blue) and the FITC (yellow) signals were taken using a thermoelectrically cooled charge coupled camera (Photometrics, Tucson, AZ) and merged using image analysis software (Boyle, 1992).

e. Sequencing of the Human PEA3 Locus

The human PEA3 locus was sequenced on both strands from a region 3.0 kb upstream of the putative transcriptional start site through all 13 exons. The sequencing was performed by either sequencing subclones of cosmid sequences using T3 and T7 primers or by using oligonucleotide primers generated from known cDNA sequence or from newly charted regions upstream of the transcriptional start site or within long intron regions.

f. Computer-Assisted Sequence Analysis and Alignment of Human and Mouse PEA3

The genomic sequences corresponding to mouse and human PEA3 were entered into the sequence editor (EditSeq) subprogram of Lasergene®, the sequence analysis

software package by DNASTAR™. For sequence alignment, the MegAlign subprogram was utilized by specifying a pair-wise alignment using the Martinez/Needleman-Wunsch algorithm. This program aligns sequences based upon criteria which specifies the minimum match required to align a stretch of sequence and penalizes itself for introducing gaps in the sequences.

Transcription factor consensus sequences were identified in both genomic mouse and human PEA3 sequence using the MatInspector® program (Quandt *et al.*, 1996). This program was accessed at the following internet address: http://transfac.gbf-braunschweig.de.welcome.html. This algorithm is unique in that nucleotide position is weighted using a matrix such that mismatches in less conserved positions are more easily tolerated than mismatches at highly conserved positions (Wingender *et al.*, 1996; Wingender *et al.*, 1997). The Transfac database, which contains greater than 1500 consensus sequences, was used to create the transcription factor consensus matrices that the MatInspector program utilizes.

RESULTS

1. PCR Amplification of the Putative Promoter Region for PEA3

Sequences flanking the putative promoter of the mouse PEA3 gene were amplified by PCR and cloned upstream of the luciferase reporter gene. The oligonucleotide primers which were designed for the amplification are listed in Table 2. The name of each construct represents the region of the PEA3 promoter which has been cloned.

To confirm the sizes of the PCR products, the amplified DNA fragments were separated on a 1% agarose gel and were compared with a 100 bp DNA ladder marker. An example of the separation of various promoter fragments is shown in Fig. 3. The PCR products for the following amplified regions migrate as ethidium bromide stained DNA bands which can be compared to the 100 bp DNA ladder (lanes 1 and 9). The regions, relative to the major transcription initiation site, +1, in lanes 2 through 8 are -356 to +21, -356 to +121, -356 to +221, -356 to +321, -356 to +421, -356 to +521, and -356 to +621, and their sizes are 377 bp, 477 bp, 577 bp, 677 bp, 777 bp, 877 bp and 977 bp, respectively. Lanes 10 through 14 contain the regions from +1 to +21, +1 to +121, +1 to +221, +1 to +321, +1 to +421 and their sizes are 21 bp, 121 bp, 221 bp, 321 bp and 421 bp, respectively.

Table 2. Oligonucleotide Primers for Amplification of the Putative Promoter of Mouse PEA3

The name of each putative PEA3 promoter/luciferase construct represents the region of the sequence which has been amplified and cloned from mouse genomic DNA. Upstream primers possess the *Bgl*II endonuclease restriction sequence. Downstream primers possess the *Hind*III restriction sequence.

CONSTRUCT NAME	PRIMER DIRECTION	SEQUENCE OF PRIMER (5' to 3')
-1341+676 *luc*	upstream	GGAAGATCTC TCATGAAAAG GGGGGTCCCT AGAGG
	downstream	CCCAGGCTTC GGGCCGCTCC CTCCGGCCAC AGG
-1226+676 *luc*	upstream	GGAAGATCTC CTTCTCTGAC TCCTACAGGC
	downstream	CCCAGGCTTC GGGCCGCTCC CTCCGGCCAC AGG
-1029+676 *luc*	upstream	GGAAGATCTC CAGCACCAGT CTGACACAGC
	downstream	CCCAGGCTTC GGGCCGCTCC CTCCGGCCAC AGG
-826+676 *luc*	upstream	GGAAGATCTG CGGGTTGAAT GATAAGCGGG TCA
	downstream	CCCAGGCTTC GGGCCGCTCC CTCCGGCCAC AGG
-756+676 *luc*	upstream	GGAAGATCTT GAGCCAGTTA AATTTACTGA G
	downstream	CCCAGGCTTC GGGCCGCTCC CTCCGGCCAC AGG
-656+676 *luc*	upstream	GGAAGATCTC CTAATGCCCC TCAGTGACAA GC
	downstream	CCCAGGCTTC GGGCCGCTCC CTCCGGCCAC AGG
-556+676 *luc*	upstream	GGAAGATCTC GCTCCGCAGC ACCACGTTATGG
	downstream	CCCAGGCTTC GGGCCGCTCC CTCCGGCCAC AGG
-456+676 *luc*	upstream	GGAAGATCTT GCCTCCCCCT CCCCCCAGGC ACA
	downstream	CCCAGGCTTC GGGCCGCTCC CTCCGGCCAC AGG
-356+676 *luc*	upstream	GGAAGATCTC CCAAGACCCA GGTTGGAACC CGTGG
	downstream	CCCAGGCTTC GGGCCGCTCC CTCCGGCCAC AGG
-256+676 *luc*	upstream	GGAAGATCTT ATTTTTTTAT GAATGGAAGT CC
	downstream	CCCAGGCTTC GGGCCGCTCC CTCCGGCCAC AGG
-100+676 *luc*	upstream	GGAAGATCTA AGTCAATGAA ACAAAGGGAA
	downstream	CCCAGGCTTC GGGCCGCTCC CTCCGGCCAC AGG
-50+676 *luc*	upstream	GGAAGATCTA ACGGAGGCCA AGGCAAAGGA
	downstream	CCCAGGCTTC GGGCCGCTCC CTCCGGCCAC AGG
+1+676 *luc*	upstream	GGAAGATCTC ACAACTGTCT GCTGCGCCCG GAAAA
	downstream	CCCAGGCTTC GGGCCGCTCC CTCCGGCCAC AGG
-1341 *luc*	upstream	GGAAGATCTC TCATGAAAAG GGGGGTCCCT AGAGG
	downstream	CCCAAGCTTA GCCGGGGGAG CAGCTGATTG GT
-826 *luc*	upstream	GGAAGATCTG CGGGTTGAAT GATAAGCGGG TCA
	downstream	CCCAAGCTTA GCCGGGGGAG CAGCTGATTG GT
-356 luc	upstream	GGAAGATCTC CCAAGACCCA GGTTGGAACC CGTGG
	downstream	CCCAAGCTTA GCCGGGGGAG CAGCTGATTG GT
-356+621 *luc*	upstream	GGAAGATCTC CCAAGACCCA GGTTGGAACC CGTGG
	downstream	CCCAAGCTTG GAGGGGGGAG GCAGTTCGCG
-356+521 *luc*	upstream	GGAAGATCTC CCAAGACCCA GGTTGGAACC CGTGG
	downstream	CCCAAGCTTT GGGACCTGCG CGGCCGAGCC
-356+421 *luc*	upstream	GGAAGATCTC CCAAGACCCA GGTTGGAACC CGTGG
	downstream	CCCAAGCTTC CTGCTCAGCC CAACTGGTTA
-356+221 *luc*	upstream	GGAAGATCTC CCAAGACCCA GGTTGGAACC CGTGG
	downstream	CCCAAGCTTG GTGTCTGGGA GATGAAGCGG
-356+121 *luc*	upstream	GGAAGATCTC CCAAGACCCA GGTTGGAACC CGTGG
	downstream	CCCAAGCTTG GATGCTCAGC ATGCCCGGGG
-356+21 *luc*	upstream	GGAAGATCTC CCAAGACCCA GGTTGGAACC CGTGG
	downstream	CCCAAGCTTC CGGGCGCAGC AGACAGTTGT
+1+621 *luc*	upstream	GGAAGATCTC ACAACTGTCT GCTGCGCCCG GAAAA
	downstream	CCCAAGCTTG GAGGGGGGAG GCAGTTCGCG
+1+521 *luc*	upstream	GGAAGATCTC ACAACTGTCT GCTGCGCCCG GAAAA
	downstream	CCCAAGCTTT GGGACCTGCG CGGCCGAGCC
+1+421 *luc*	upstream	GGAAGATCTC ACAACTGTCT GCTGCGCCCG GAAAA
	downstream	CCCAAGCTTC CTGCTCAGCC CAACTGGTTA
+1+221 *luc*	upstream	GGAAGATCTC ACAACTGTCT GCTGCGCCCG GAAAA
	downstream	CCCAAGCTTG GTGTCTGGGA GATGAAGCGG
+1+121 *luc*	upstream	GGAAGATCTC ACAACTGTCT GCTGCGCCCG GAAAA
	downstream	CCCAAGCTTG GATGCTCAGC ATGCCCGGGG
+1+21 *luc*	upstream	GGAAGATCTC ACAACTGTCT GCTGCGCCCG GAAAA
	downstream	CCCAAGCTTC CGGGCGCAGC AGACAGTTGT

Figure 3. Example of PCR Products Separated by Agarose Gel Electrophoresis

PCR products (0.1 volumes of the reaction) were separated by agarose gel (1%) electrophoresis at 90V for 2 hours. Lanes 1 and 9 represent the 100 bp DNA ladder marker. Lanes 2 - 8 and 10 - 14 represent the amplified sequences from -356 to +21, -356 to +121, -356 to +221, -356 to +321, -356 to +421, -356 to +521,-356 to +621, +1 to +21, +1 to +121, +1 to +221, +1 to +321, and +1 to +421, respectively. Lane 15 represents the negative control PCR reaction in which no DNA template was added.

bp
100
600
1600
2000
100 bp Ladder
-356+21
-356+121
-356+221
-356+321
-356+421
-356+521
-356+621
100 bp Ladder
+1+21
+1+121
+1+221
+1+321
+1+421
NO TEMPLATE

a. Restriction Endonuclease and Sequencing Analyses

Restriction endonuclease analysis followed by sequencing was used to confirm that the anticipated fragments were cloned in the correct orientation upstream of the luciferase reporter gene. Restriction endonuclease analysis of pGL3 plasmid DNA isolated from the transformants showed that the expected PCR amplified promoter region was inserted into the reporter vector. An example of the restriction endonuclease analyses of the promoter constructs is shown in Fig. 4. In lane 2, the promoterless plasmid, pGL3 basic (PL *luc*), has been linearized with *Bgl*II and *Hin*dIII endonucleases. Lanes 3 through 7, 9 and 10, are the regions from -1341 to +676, -1029 to +676, -826 to +676, -656 to +676, -456 to +676, -356 to +676, -256 to +676, which were excised from the pGL3 reporter vector, respectively. Estimation of the sizes of the DNA fragments was by comparison of ethidium bromide stained DNA ladder markers separated by agarose gel electrophoresis.

Sequence analysis of each pGL3 chimeric plasmid was by the Sanger method. Two primers designed for the pGL3 plasmid were used in separate analyses to obtain the sequence of the insert in both directions. To sequence longer promoter regions internally, primers that were originally used to generate deletions of these regions were used as sequencing primers. Sequencing of the pGL3/PEA3 chimera verified that the correct promoter sequence had been cloned in the correct orientation.

The regions of the PEA3 locus that were correctly amplified and cloned into the luciferase reporter vectors are shown in Fig. 5. Deletions were constructed from both the 5' end (Fig. 5a.) or the 3' end (Fig. 5b.) of the putative promoter region.

Figure 4. Example of Restriction Endonuclease Analysis of PEA3 Promoter/Luciferase Constructs

Purified plasmid DNA (0.5 μg) from transformants was digested with *Bgl*II and *Hin*dIII endonucleases and then subjected to agarose gel (1%) electrophoresis at 90V for one hour to confirm the correct size of the cloned DNA fragment. Lanes 1 and 8 are 100 bp DNA ladder marker, respectively. Lanes 2 - 7 and 9 - 10 are the promoterless vector (PL *luc*), -1341+676 *luc*, -1029+676 *luc*, -826+676 *luc*, -656+676 *luc*, -456+676 *luc*, -356+676 *luc*, and -256+676 *luc*, respectively.

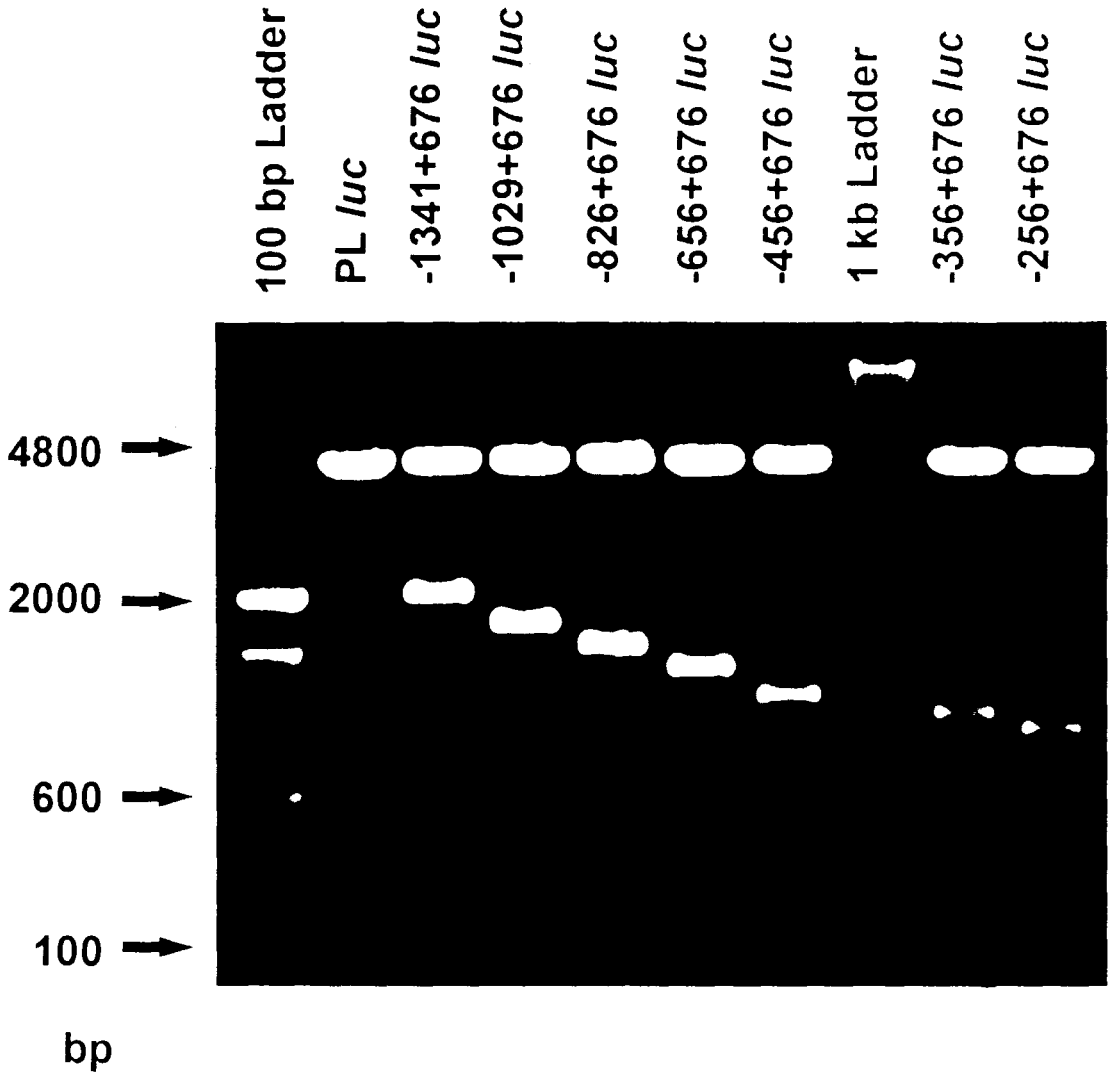
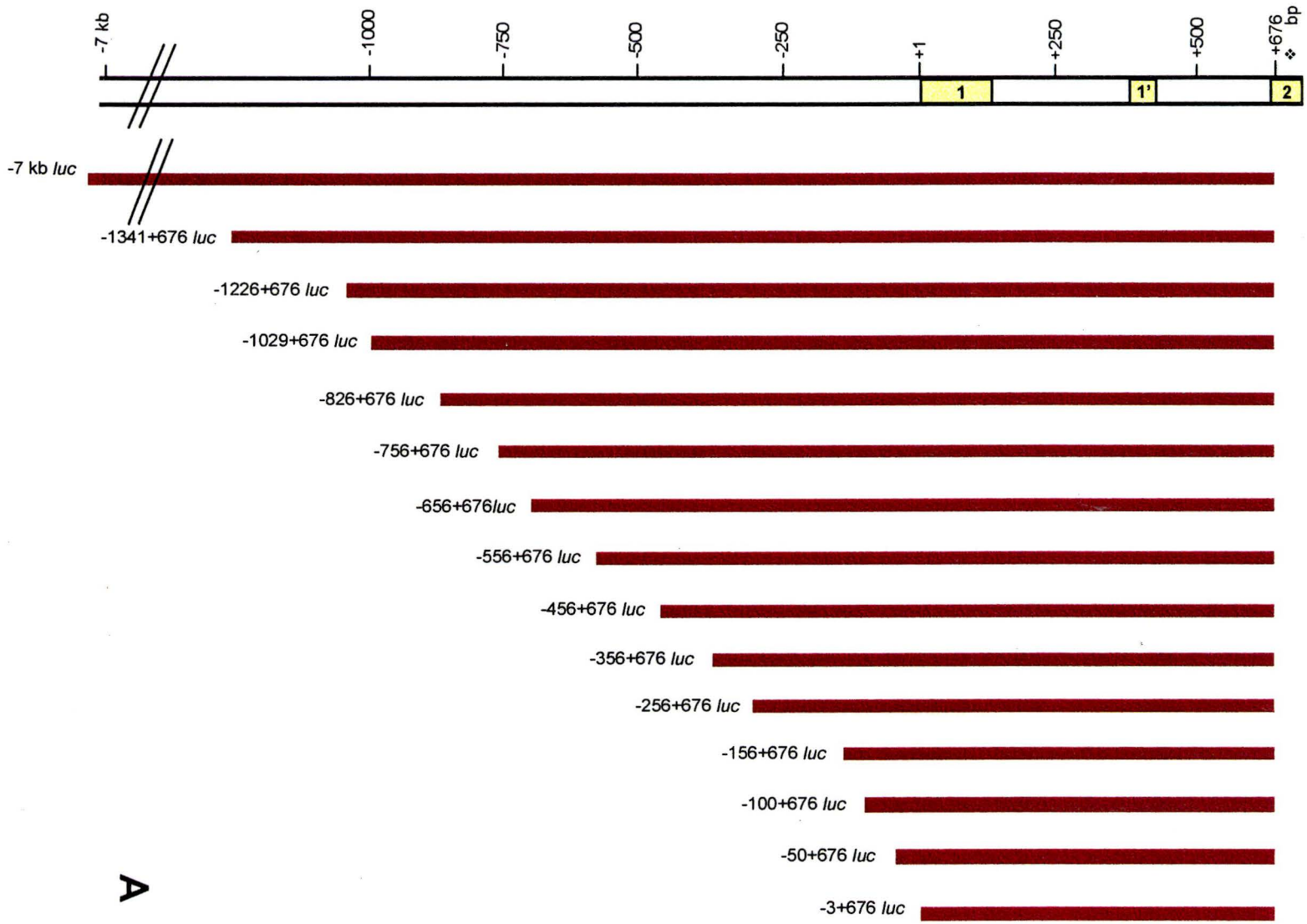

100 bp Ladder
PL *luc*
-1341+676 *luc*
-1029+676 *luc*
-826+676 *luc*
-656+676 *luc*
-456+676 *luc*
1 kb Ladder
-356+676 *luc*
-256+676 *luc*
4800
2000
600
100
bp

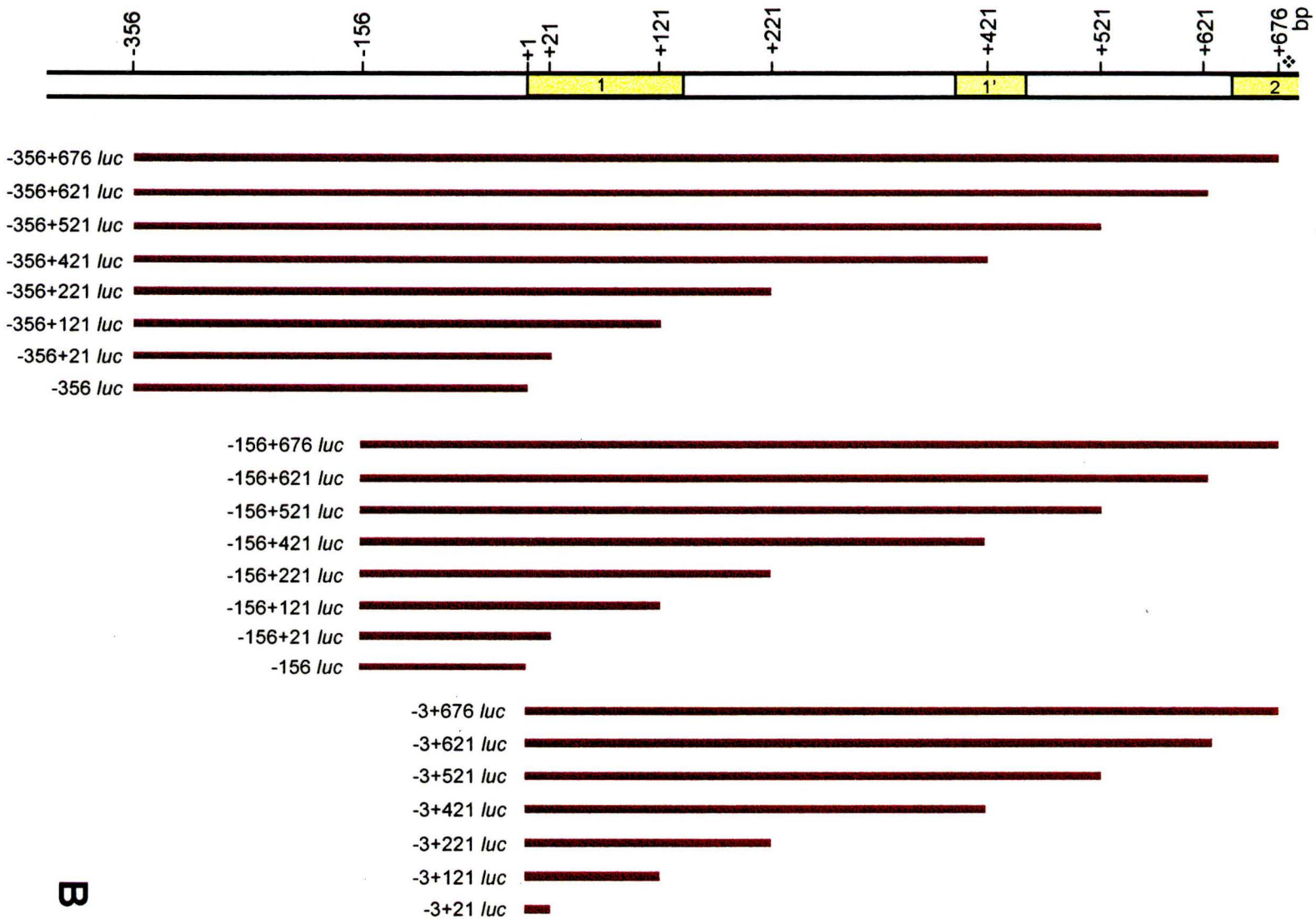
Figure 5. Regions of the PEA3 Promoter Cloned into the pGL3 Reporter Vector

The numbers above the bar represent the nucleotide position relative to the major transcription start site at exon 1. Exons are indicated by yellow boxes. The putative translation start codon (+677) is indicated by a diamond. The region of the promoter contained in each construct is indicated by a bar with the name of each construct shown on the left.

A. 5' Deletions of the Putative Promoter
B. 3' Deletions of the Putative Promoter


-7 kb
-1000
-750
-500
-250
+1
+250
+500
+676 bp
1
1'
2
-7 kb luc
-1341+676 luc
-1226+676 luc
-1029+676 luc
-826+676 luc
-756+676 luc
-656+676luc
-556+676 luc
-456+676 luc
-356+676 luc
-256+676 luc
-156+676 luc
-100+676 luc
-50+676 luc
-3+676 luc

A


-356
-156
+1
+21
+121
+221
+421
+521
+621
+676
bp
1
1'
2
-356+676 luc
-356+621 luc
-356+521 luc
-356+421 luc
-356+221 luc
-356+121 luc
-356+21 luc
-356 luc
-156+676 luc
-156+621 luc
-156+521 luc
-156+421 luc
-156+221 luc
-156+121 luc
-156+21 luc
-156 luc
-3+676 luc
-3+621 luc
-3+521 luc
-3+421 luc
-3+221 luc
-3+121 luc
-3+21 luc

B

2. Luciferase Assays

The luciferase assay was used to assess the relative *in vivo* promoter activity of various deletions of the putative PEA3 promoter. The luciferase activity of the SV40 early promoter-luciferase construct, SV *luc*, was set at 100 luciferase units for each assay to allow comparison of each putative promoter clone. As a control for transfection efficiency, pSV-βgal was co-transfected with each putative PEA3 promoter construct. Relative luciferase activity, including the standard error of the mean (SEM), for each promoter clone was calculated. The experiments were performed three times using two or more different preparations of DNA. Each transfection was performed in duplicate and each cell lysate was assayed for luciferase activity and β-galactosidase activity in duplicate.

a. Promoter Efficiency of 5' Deletions of the Putative Promoter

Relative luciferase activity of the 5' deletions of the putative promoter regions are presented, along with the SEM, in Fig. 6a. The luciferase reporter construct containing the longest putative promoter region, -7 kb *luc*, had an activity of 7.6 ± 1.2 relative to SV *luc*. In this same experiment, the promoterless construct, PL *luc*, had an activity of 5.5 ± 1.2. The values are not significantly different based on the analysis of the variance of the two sets of data, indicating that this DNA fragment did not possess the ability to direct transcription of the reporter gene above a basal level. The reporter construct bearing the sequence from -1341 bp to +676 bp of the mouse PEA3 gene had a relative mean activity of 11.0 ± 1.4 . This value is significantly greater than the value for the basal promoter and may imply that a silencer element had been removed between -1341 and -7000. The

Figure 6. Luciferase Activity of PEA3 Promoter/Luciferase Constructs

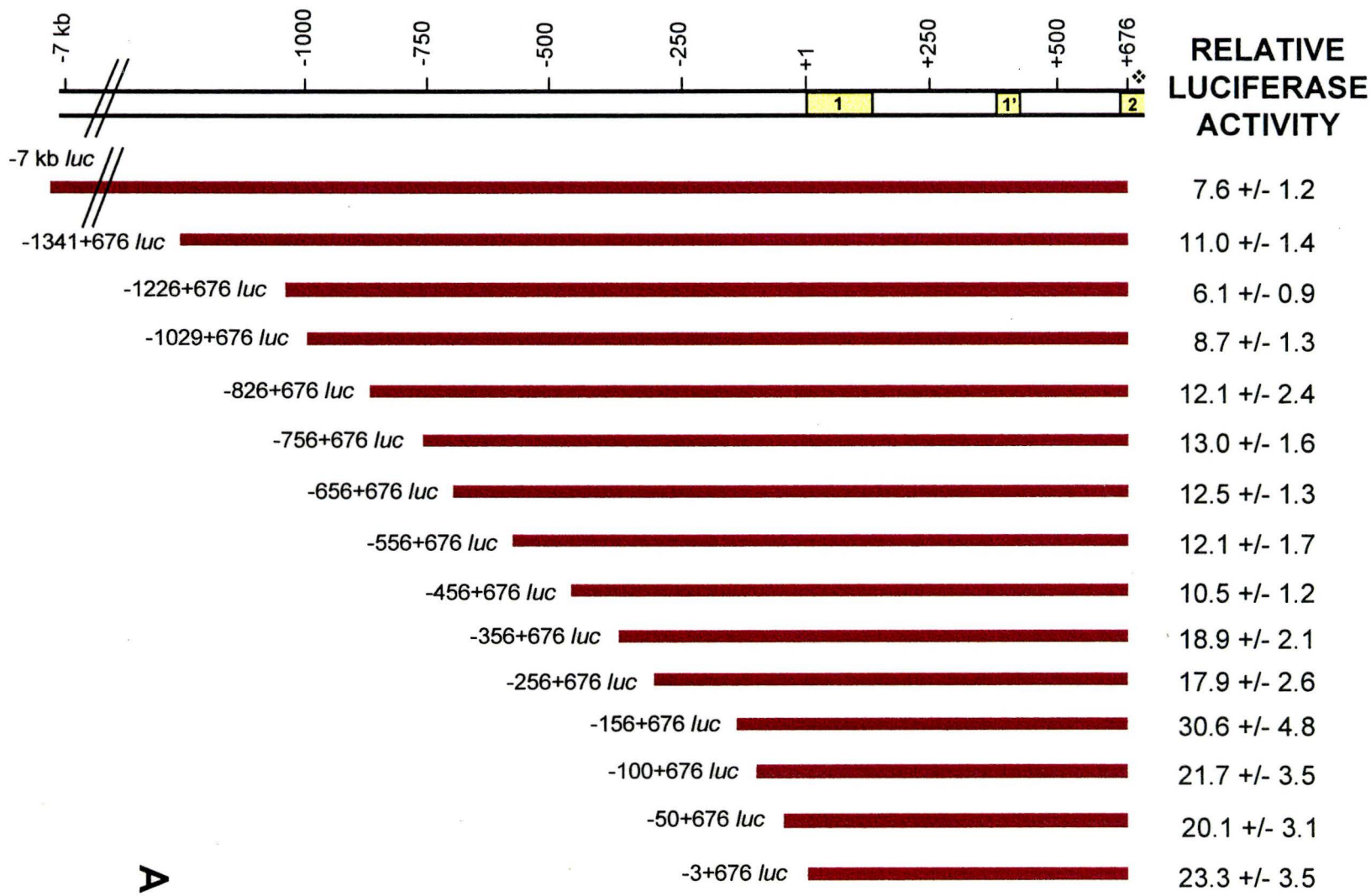
The mean relative luciferase activity of each PEA3 promoter construct is presented along with the standard error of the mean (SEM). Luciferase activity of each construct in FM3A cells is calculated relative to the luciferase activity of SV *luc* (SV40 early promoter driving luciferase expression) and is the average of 3 experiments of which each transfection and assay is performed in duplicate. Transfection efficiency is internally controlled for with the pSV-βgal reporter plasmid.

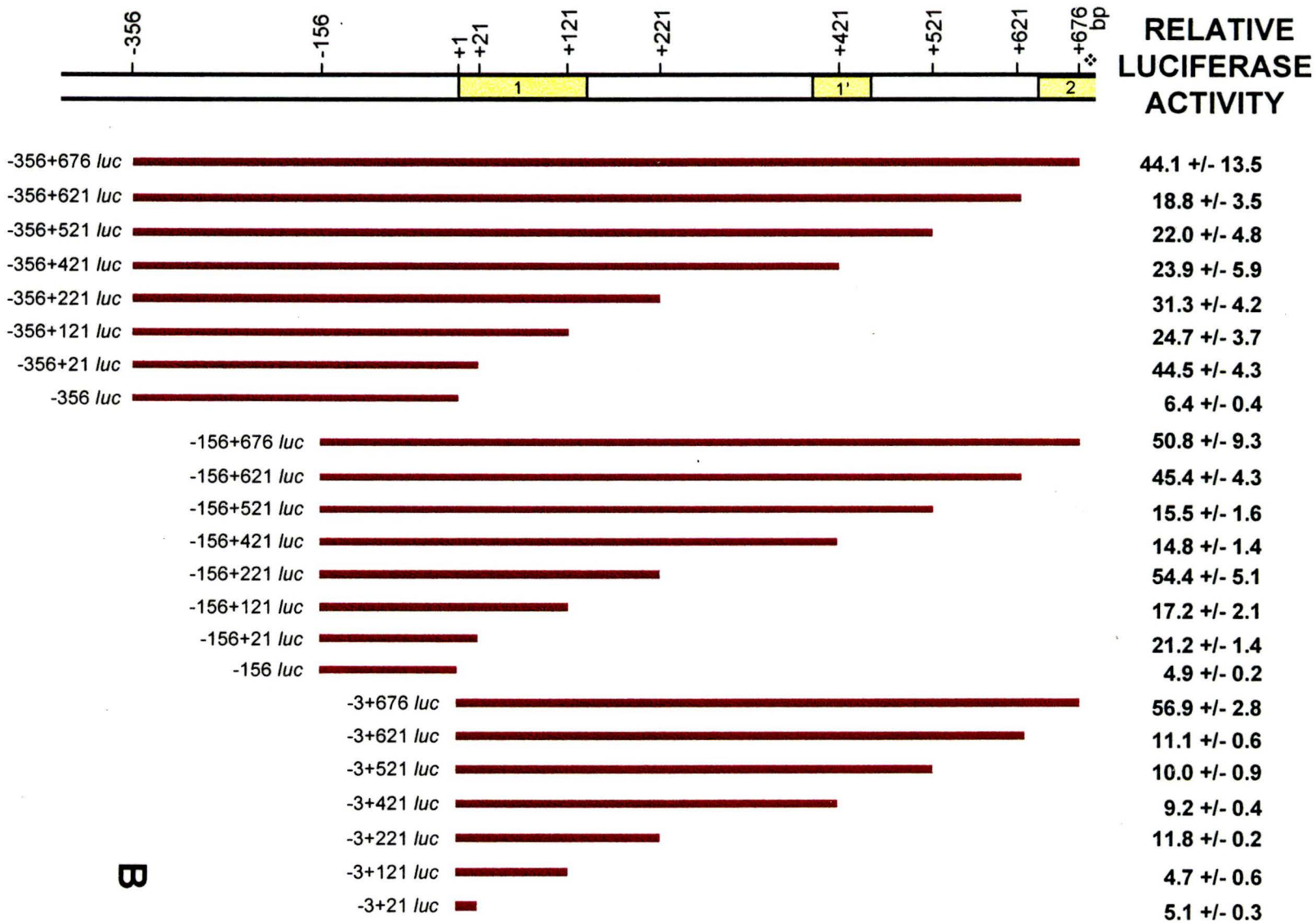
A. 5' Deletions of the Putative Promoter

Relative luciferase activity for the promoterless vector, PL *luc*, was 5.4 ± 1.2 in FM3A cells.

B. 3' Deletions of the Putative Promoter

Relative luciferase activity for the promoterless vector, PL *luc*, was 4.3 ± 0.5 in FM3A cells.


-7 kb
-1000
-750
-500
-250
+1
+250
+500
+676
1
1'
2
RELATIVE LUCIFERASE ACTIVITY
-7 kb luc
7.6 +/- 1.2
-1341+676 luc
11.0 +/- 1.4
-1226+676 luc
6.1 +/- 0.9
-1029+676 luc
8.7 +/- 1.3
-826+676 luc
12.1 +/- 2.4
-756+676 luc
13.0 +/- 1.6
-656+676 luc
12.5 +/- 1.3
-556+676 luc
12.1 +/- 1.7
-456+676 luc
10.5 +/- 1.2
-356+676 luc
18.9 +/- 2.1
-256+676 luc
17.9 +/- 2.6
-156+676 luc
30.6 +/- 4.8
-100+676 luc
21.7 +/- 3.5
-50+676 luc
20.1 +/- 3.1
-3+676 luc
23.3 +/- 3.5
A


-356
-156
+1
+21
+121
+221
+421
+521
+621
+676 bp
1
1'
2
RELATIVE LUCIFERASE ACTIVITY
-356+676 luc
44.1 +/- 13.5
-356+621 luc
18.8 +/- 3.5
-356+521 luc
22.0 +/- 4.8
-356+421 luc
23.9 +/- 5.9
-356+221 luc
31.3 +/- 4.2
-356+121 luc
24.7 +/- 3.7
-356+21 luc
44.5 +/- 4.3
-356 luc
6.4 +/- 0.4
-156+676 luc
50.8 +/- 9.3
-156+621 luc
45.4 +/- 4.3
-156+521 luc
15.5 +/- 1.6
-156+421 luc
14.8 +/- 1.4
-156+221 luc
54.4 +/- 5.1
-156+121 luc
17.2 +/- 2.1
-156+21 luc
21.2 +/- 1.4
-156 luc
4.9 +/- 0.2
-3+676 luc
56.9 +/- 2.8
-3+621 luc
11.1 +/- 0.6
-3+521 luc
10.0 +/- 0.9
-3+421 luc
9.2 +/- 0.4
-3+221 luc
11.8 +/- 0.2
-3+121 luc
4.7 +/- 0.6
-3+21 luc
5.1 +/- 0.3
B

construct containing the region from -1226 to +676 had an activity of 6.1 ± 1.9, a value which is not significantly different from the promoterless construct. The relative mean luciferase activity for the constructs with 5' endpoints at -1029, -826, and -756 were 8.7 ± 1.3, 12.1 ± 2.4 and 13.0 ± 1.6, respectively. These results reveal that deletions from the 5' end increased reporter gene expression. The luciferase activity values for -656+676 *luc*, -556+676 *luc*, and -456+676 *luc* were 12.5 ± 1.3, 12.1 ± 1.7 and 10.5 ± 1.2, respectively, which are not significantly different from each other based on the analysis of the variance of each set of data. The values obtained for -356+676 *luc* and -256+676 *luc* were 18.9 ± 2.1 and 17.9 ± 2.6, respectively. Although these two constructs have activities which are not significantly different from each other, both are, in fact, significantly greater in activity than -456+676 *luc*. In these experiments, the promoter fragment bearing the region from -156 to +676 had the greatest relative luciferase activity of 30.6 ± 4.8. This value is significantly greater than the -256+676 *luc* construct, implicating that still further deletions of the 5' end increased promoter activity. The luciferase activity measured for -100+676 *luc*, -50+676 *luc* and -3+676 *luc* was 21.7 ± 3.5, 20.1 ± 3.1 and 23.3 ± 3.5, respectively. These values are not significantly different from one another, nor are any of them significantly different from the construct containing the region from -156 to +676. These results imply that deletions from -156 to -3 did not significantly affect luciferase activity.

b. Promoter Efficiency of 3' Deletions of the Putative Promoter

To analyze the effect of the sequences between -3 and +676 on promoter activity, luciferase reporter constucts were designed which removed 3' sequences from the putative

promoter region of PEA3. Relative luciferase activity of the 3' deletions of the putative promoter regions are presented, along with the SEM, in Fig. 6b.

Three clones from the series of 5' deletions: -356+676, -156+676, and -3+676, were selected and deletion constructs of each were made from the 3' end and then subjected to analysis. For the series containing the 5' endpoint at -356, the first deletion which removes 55 bp was accompanied by a statistically significant decrease in luciferase activity: the activity of -356+676 was 44.1 ± 13.5 and the activity of -356+621 was 18.8 ± 3.5. Further 3' deletions created -356+521*luc*, -356+421*luc* , -356+221*luc*, and -356+121 *luc* which had relative luciferase activity values of 22.0 ± 4.8, 23.9 ± 5.9, 31.3 ± 4.2 and 24.7 ± 3.7. These values are neither significantly different from each other nor the construct containing sequences form -356 to +621 based on the analysis of the variance within each set of data. The above set of results show that removal of sequences from +676 to +621 decreased luciferase activity slightly over two-fold and a further deletion to +121 did not have an observable effect. The removal of the sequence from +121 to +21, creating -356+21 *luc*, appeared to restore the activity to that of the full-length construct: 44.5 ± 4.3. The deletion of the sequences from +21 to -1 resulted in a statistically significant decrease to 6.4 ± 0.4. This value is not statistically different from the activity promoterless construct, PL *luc* which, in this experiment, had an activity of 4.3 ± 0.5. The activity of -356 *luc*, whose 3' endpoint is at -1 relative the major transcriptional initiation site, appears to be consistent with the removal of the putative Inr (initiator) element of promoters.

The series of 3' deletions which have their 5' endpoint at -156 were also analyzed for promoter activity following successive deletions. The construct containing -156 to +676 shows a relatively high level of promoter activity at a value of 50.8 ± 9.3 luciferase units. Unlike the construct with the -356 endpoint, of which its activity decreased upon removal of the region from +676 to +621, there is no significant difference in the values obtained for both -156+676 *luc* and -156+621 *luc* (50.8 ± 9.3 and 45.4 ± 4.3, respectively). There was a three fold decrease in luciferase activity when 100 bp was removed from -156+621 *luc*, thus the construct bearing the sequence from -156 to +521 had an activity of 15.5 ± 1.6 relative luciferase units. A further deletion of 100 bp gave rise to a value of 14.8 ± 1.4 which is not statistically different from the activity of -156+521 *luc*. The promoter construct containing the region from -156 to +221 *luc* had an activity of 54.4 ± 5.1. Further 3' deletions created -156+121 *luc* and -156 +21 *luc* with activities of 17.2 ± 2.1 and 21.2 ± 1.4, respectively. Although these two constructs have activities which are not significantly different from each other, both are, in fact, significantly lower in activity than -156+221 *luc*. As observed above, disruption of the sequences surrounding the major transcription initiation site, showed that the construct containing the sequence from -156 to -1 (-156 *luc*) had an activity of 4.9 ± 0.2 which was not significantly different from the activity of the promoterless construct.

Evidence that the genomic sequence containing the first exon through to the second exon possess sequences with promoter activity was the high amount of promoter activity from the construct that contains the region from -3 to +676 bp. This luciferase construct

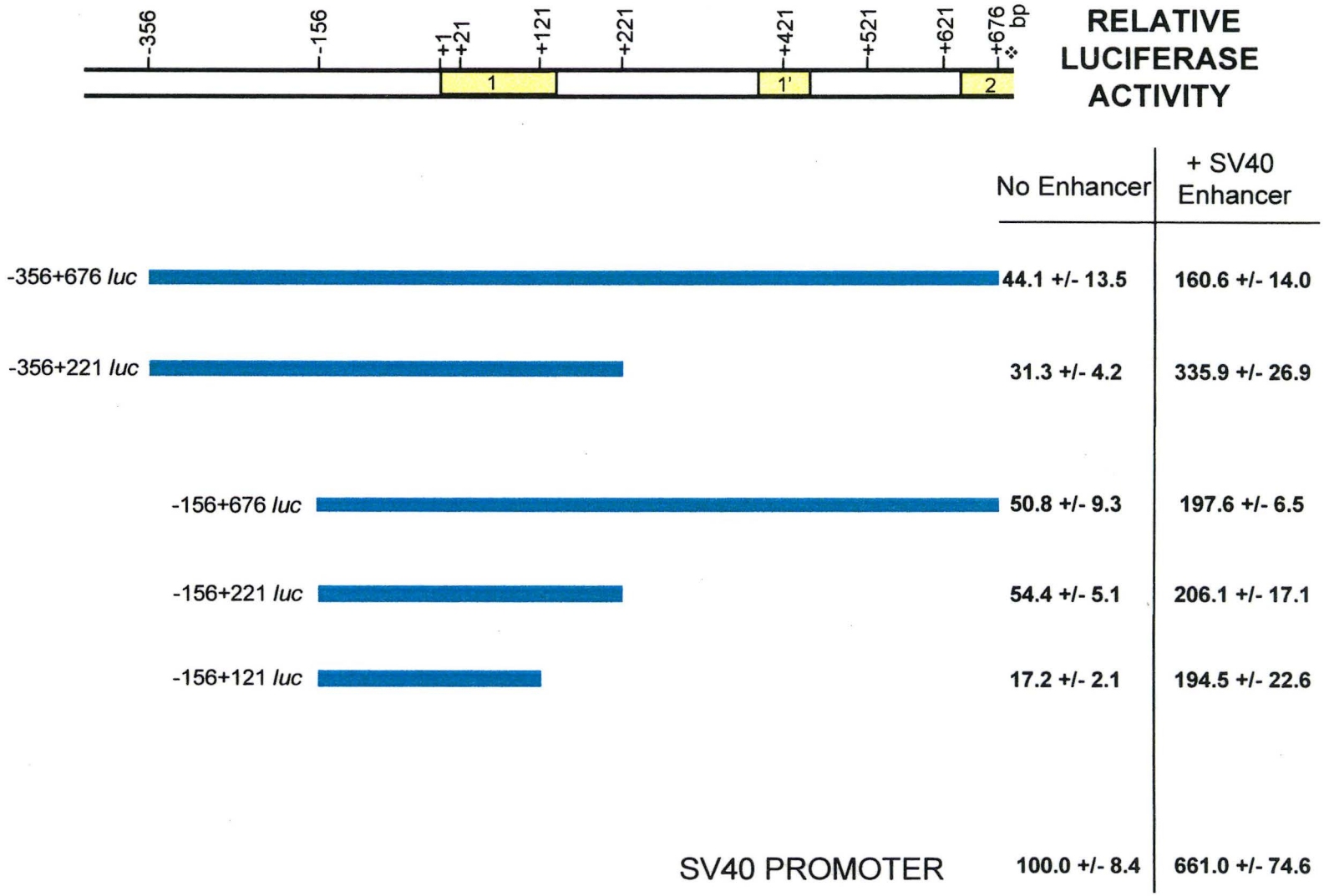
had a relative activity of 56.9 ± 2.8 luciferase units. Deletions from the 3' end of this construct produced the following putative promoter clones, -3+621 *luc*, -3+521 *luc* and -3+421 *luc*, that had activities of 11.1 ± 0.6, 10.0 ± 0.9 and 9.2 ± 0.4, respectively. Based on the analysis of the variance within the three sets of data, these three constructs have activities which are not significantly different from each other, however, all three are significantly lower in activity than -3+676 *luc*. A further 3' deletion of 200 bp from -3+421 *luc* gives rise to -3+221 *luc* which had a relative activity of 11.8 ± 0.2. The smallest putative promoter fragments which were assayed for promoter acitivity were -3+121 *luc* and -3+21 *luc*. These two constructs had activities of 4.7 ± 0.6 and 5.1 ± 0.3, respectively, which are not statistically different from the activity of the promoterless construct, PL *luc*.

c. Assessment of SV40 Enhancer-Containing Constructs

In order to verify that the putative promoter region of PEA3 was a *bona fide* promoter, the SV40 enhancer element was cloned at a distance upstream of the putative promoter to determine if luciferase expression was enhanced. The results of the luciferase assays of these enhancer-containing constructs are presented in Fig. 7. In these experiments, the construct which contained the largest region of mouse genomic PEA3 was -356+676 *luc*. In the presence of the SV40 enhancer element, the luciferase activity was enhanced 3.6 fold, from 44.1 ± 13.5 to 160.6 ± 14.0 relative luciferase units. In the presence of the SV40 enhancer, the activity of the -356+221 *luc* increased 10.7 fold, from 31.3 ± 4.2 to 335.9 ± 26.9 luciferase units. The putative PEA3 promoter construct containing the region from -156 to +676 had an activity which increased 3.8 fold (from

Figure 7. Luciferase Activity of PEA3 Promoter/Luciferase /SV40 Enhancer Constructs

The mean relative luciferase activity of PEA3 promoter constucts is presented along with the standard error of the mean (SEM). The SV40 enhancer was cloned upstream of the regions which are shown. Luciferase activity of each construct is calculated relative to the luciferase activity of SV *luc* (SV40 early promoter driving luciferase expression) and is the average of three experiments of which each transfection and assay is performed in duplicate. Transfection efficiency is internally controlled for with the pSV-βgal reporter plasmid. The luciferase activity of the promoter constructs, containing the enhancer sequence are shown on the right. The luciferase activity in the absence of the enhancer is shown on the left. The relative activity of the promoterless construct containing the enhancer was 31.9 ± 5.3. In the absence of the enhancer, the activity was 4.3 ± 0.5.


-356
-156
+1
+21
+121
+221
+421
+521
+621
+676
bp
1
1'
2
RELATIVE LUCIFERASE ACTIVITY
No Enhancer
+ SV40 Enhancer
-356+676 luc
44.1 +/- 13.5
160.6 +/- 14.0
-356+221 luc
31.3 +/- 4.2
335.9 +/- 26.9
-156+676 luc
50.8 +/- 9.3
197.6 +/- 6.5
-156+221 luc
54.4 +/- 5.1
206.1 +/- 17.1
-156+121 luc
17.2 +/- 2.1
194.5 +/- 22.6
SV40 PROMOTER
100.0 +/- 8.4
661.0 +/- 74.6

50.8 ± 9.3 to 197.6 ± 6.5) in the presence of the SV40 enhancer element. The SV40 enhancer increased the activity of -156+221 *luc* 3.8 fold from 54.4 ± 5.1 to 206.1 ± 17.1 luciferase units. The SV40 enhancer had the greatest effect on -156+676 *luc*. The activity of this construct was increased 11.3 fold from 17.2 ± 2.1 to 194.5 ± 22.6 luciferase units. In this experiment, the SV40 enhancer was shown to enhance the SV40 early promoter 6.61 fold. It was also found to enhance the activity of the promoterless construct 7.4 fold.

3. Isolation and Characterization of Human PEA3 Cosmids

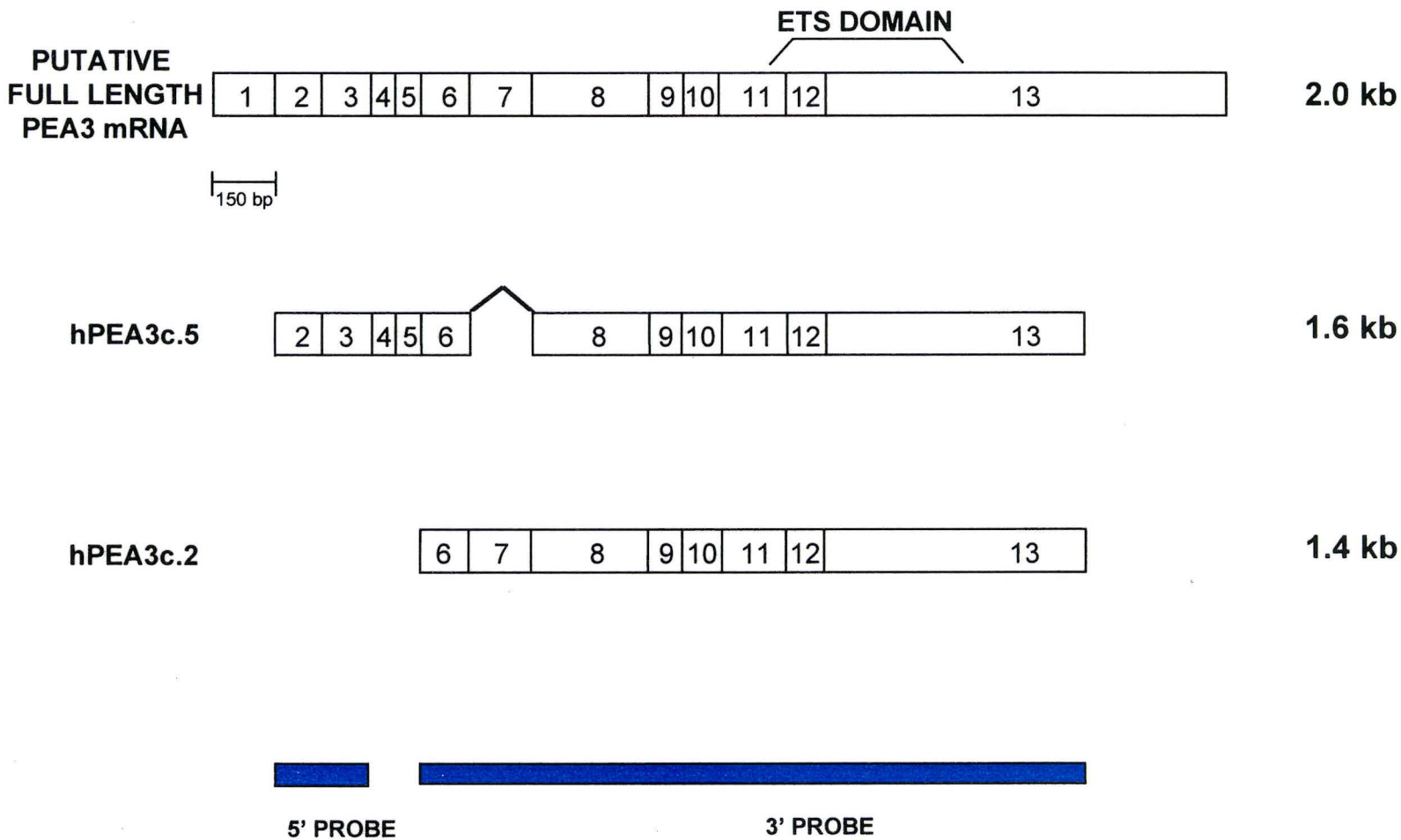
As an independent attempt to determine important regulatory elements in the PEA3 promoter, the putative promoter of human PEA3 was isolated and sequenced. The basis for this study was to explore the functional significance of conserved sequences in the non-coding region of the locus.

a. Human Cosmid Library Screening

A human fetal brain genomic cosmid library was screened using two PEA3 cDNA probes (Fig. 8). The 5' probe was 225 bp and corresponded to exons 2 and 3, inclusively, of the hPEA3c.5 cDNA. The 3' probe was 1.4 kb and encompassed exons 6 through 13, inclusively, of the hPEA3c.2 cDNA. The 3' probe allowed the isolation of two cosmids, hPEA3g.4 and hPEA3g.7. Rescreening of the library with the 5' probe allowed for six more cosmids to be isolated: hPEA3g.1, hPEA3g.2, hPEA3g.3, hPEA3g.5, hPEA3g.6, and hPEA3g.8.

Figure 8. Human PEA3 cDNA Probes

The top diagram indicates the putative full length PEA3 mRNA. The coding region for the ETS domain is indicated (exons 11 through 13). The two cDNA isolates, hPEA3c.2 and hPEA3c.5, were isolated from a human placental cDNA library by Paul Lachance. The two probes generated from these cDNAs are shown at the bottom of the figure: 5' probe and 3' probe.


ETS DOMAIN
PUTATIVE FULL LENGTH PEA3 mRNA
1 2 3 4 5 6 7 8 9 10 11 12 13
2.0 kb
150 bp
hPEA3c.5
2 3 4 5 6 8 9 10 11 12 13
1.6 kb
hPEA3c.2
6 7 8 9 10 11 12 13
1.4 kb
5' PROBE
3' PROBE

b. Restriction Endonuclease and Southern Blot Analyses of Human PEA3 Cosmids

In order to characterize the cosmid isolates, restriction endonuclease and Southern analyses were performed. By digesting each cosmid isolate with restriction endonucleases, each was found to contain a minimum of 35 kb of genomic DNA (data not shown). The cosmids, hPEA3g.4 and hPEA3g.7, were digested to completion with *Eco*RI and *Xba*I endonucleases and separated by agarose gel (0.7%) electrophoresis. The DNA fragments migrated as ethidium bromide stained bands and the banding pattern was identical for both cosmids (data not shown). In order to determine which fragments shared DNA sequence with the human cDNA. The DNA fragments were immobilized on a nylon filter and hybridized to the 3' probe (Fig. 9a). The results indicate that the same DNA fragments from both hPEA3g.4 and hPEA3g.7 were capable of hybridizing to this probe and these fragments were a 8.5 kb *Xba*I fragment, a 1.6 kb and a 1.0 kb *Eco*RI fragment. The remaining cosmids were digested to completion with *Eco*RI and *Bst*XI and separated by agarose gel (0.7%) electrophoresis. The cosmids hPEA3g.1, hPEA3g.5, hPEA3g.6 were found to have an identical DNA fragment separation pattern whereas hPEA3g.2, hPEA3g.3, and hPEA3g.8 had a separation pattern of which not any were entirely identical, however, some bands coexisted in each cosmid (data not shown) implying that these isolates shared overlapping sequences. For further analysis, only one of each representative cosmid was subjected to analysis.

In order to determine which regions of these cosmids shared sequence with the human cDNA for PEA3, Southern blots were performed with both the 5' probe and the 3' probe (Fig. 9b and 9c). Hybridization of the 5' probe, which encompassed exons 2 and 3

Figure 9. Southern Blot Analyses of Human Genomic Cosmid Isolates

A. Hybridization of a 3' Probe to hPEA3g.4 and hPEA3g.7

Purified cosmid DNA was digested with *Eco*RI and *Xba*I endonucleases, separated on a 0.7% agarose gel and transferred to a nylon filter. The 3' probe (Fig. 9) was labelled with [α-^{32}P]dCTP and hybridized to the immobilized DNA fragments. Lane 1, 1 kb ladder; lanes 2 - 3, *Eco*RI digestion of hPEA3g.4 and hPEA3g.7, respectively; lanes 4 - 5, *Xba*I digestion of hPEA3g.4 and hPEA3g.7, respectively; lane 6, hPEA3c.2.

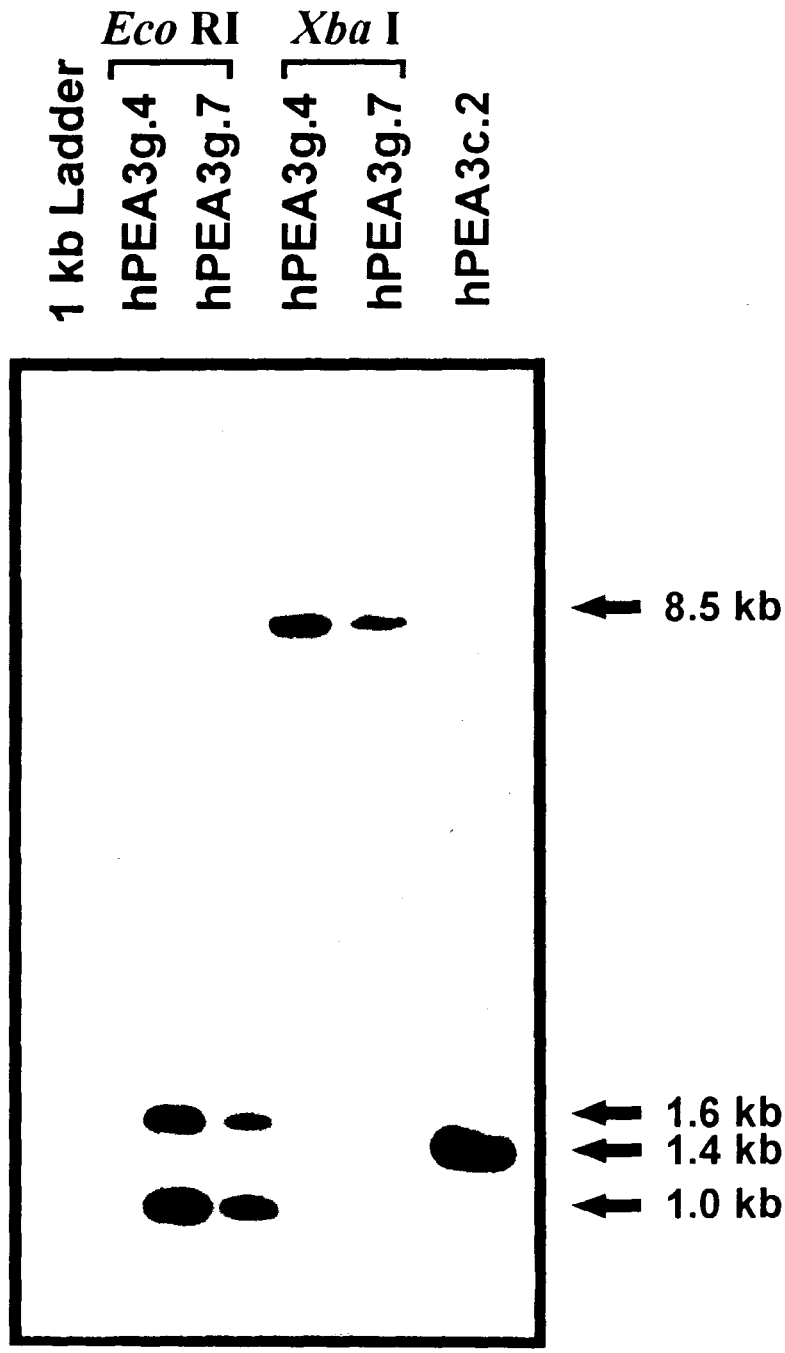
B. Hybridization of a 5' Probe to Cosmids Overlapping the 5'End of the PEA3 Gene

Purified cosmid DNA was digested with *Eco*RI and *Bst*XI endonucleases, separated on a 0.7% agarose gel and transferred to a nylon filter. The 5' probe (Fig. 9) was labelled with [^{32}P]dCTP and hybridized to the immobilized DNA fragments. Lanes 1 - 4, *Eco*RI digestion of hPEA3g.1, hPEA3g.2, hPEA3g.3, hPEA3g.8, respectively; lane 5, hPEA3c.5 (exon 2-3), lane 6, hPEA3c.2; lane 7, 1 kb ladder; lanes 8 - 11, *Bst*XI digestion of hPEA3g.1, hPEA3g.2, hPEA3g.3, hPEA3g.8, respectively.

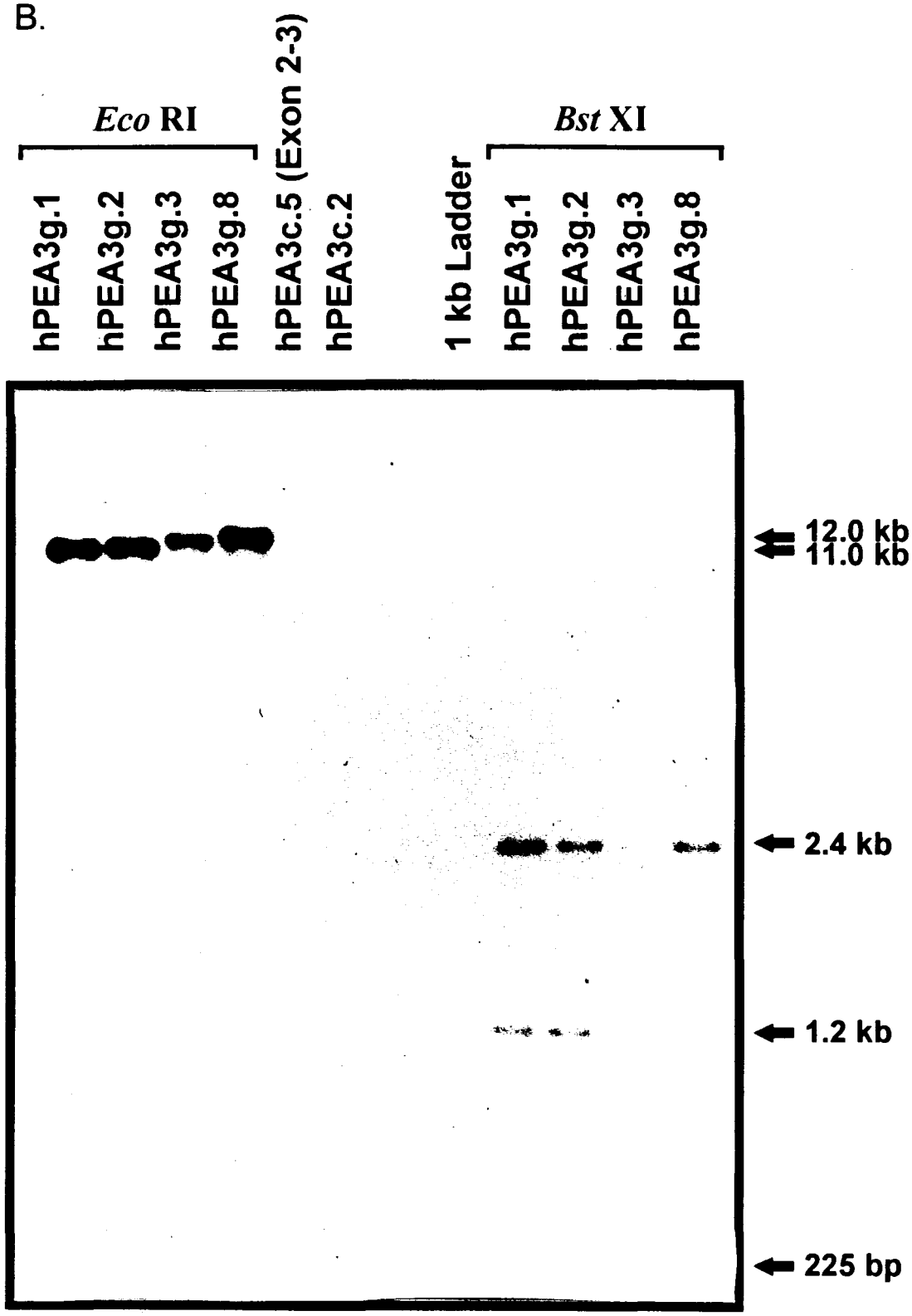
C. Hybridization of a 3' Probe to Cosmids Overlapping the 3' End of the PEA3 Gene

As above, with 3' probe (Fig. 9).

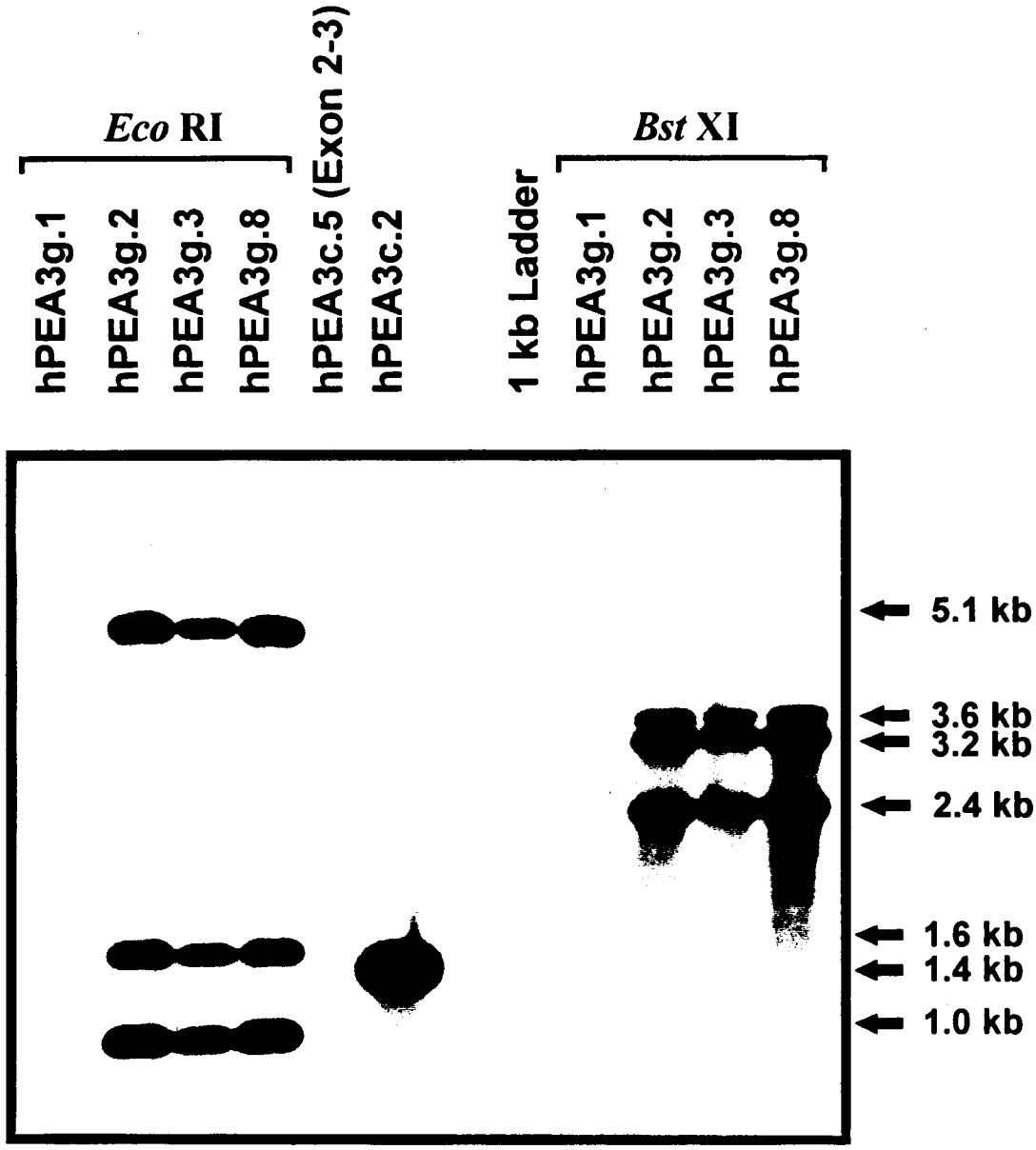
A.


1 kb Ladder
Eco RI
hPEA3g.4
hPEA3g.7
Xba I
hPEA3g.4
hPEA3g.7
hPEA3c.2
8.5 kb
1.6 kb
1.4 kb
1.0 kb

B.


Eco RI
hPEA3g.1
hPEA3g.2
hPEA3g.3
hPEA3g.8
hPEA3c.5 (Exon 2-3)
hPEA3c.2
1 kb Ladder
Bst XI
hPEA3g.1
hPEA3g.2
hPEA3g.3
hPEA3g.8
12.0 kb
11.0 kb
2.4 kb
1.2 kb
225 bp

C.


Eco RI
hPEA3g.1
hPEA3g.2
hPEA3g.3
hPEA3g.8
hPEA3c.5 (Exon 2-3)
hPEA3c.2
1 kb Ladder
Bst XI
hPEA3g.1
hPEA3g.2
hPEA3g.3
hPEA3g.8
5.1 kb
3.6 kb
3.2 kb
2.4 kb
1.6 kb
1.4 kb
1.0 kb

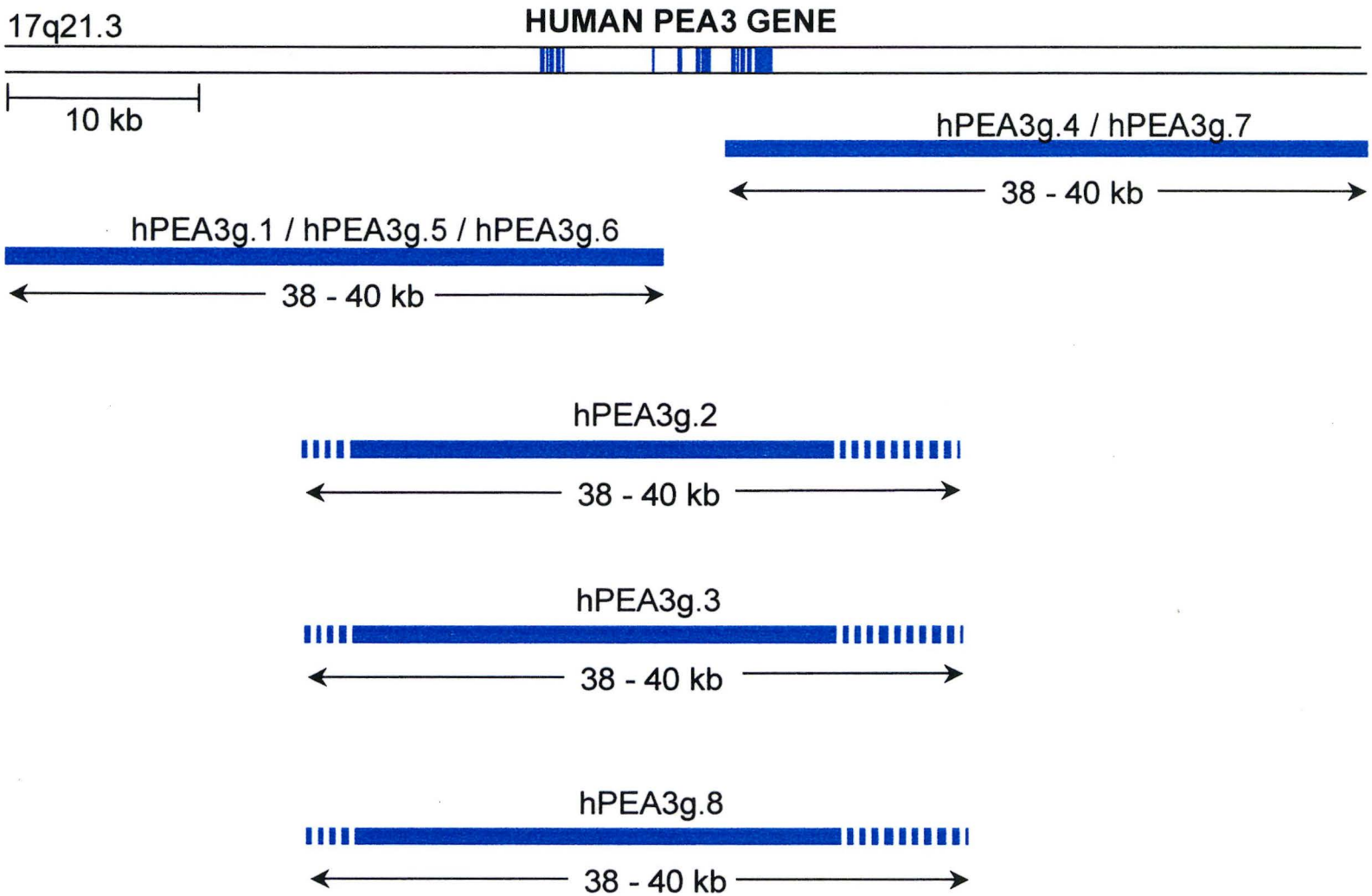
of human PEA3 cDNA, inclusively, revealed that each of the cosmids shared sequence with the 5' end of the human cDNA. For each of the cosmids hPEA3g.1, hPEA3g.2, hPEA3g.3 and hPEA3g.8, the fragments sharing sequence were an *Eco*RI fragment of ≥ 11.0 kb, one 2.4 kb and one 1.2 kb *Bst*XI fragment. Hybridization of the 3' probe, containing exons 6-13, inclusively, of human PEA3 cDNA identified 3 *Eco*RI fragments of lengths 5.1 kb, 1.6 kb and 1.0 kb and 3 *Bst*XI fragments of lengths 3.6 kb, 3.2 kb and 2.4 kb which share sequences with this region of the cDNA. The 3' probe hybridized to hPEA3g.2, hPEA3g.3 and hPEA3g.8 but not to hPEA3g.1. From the information obtained through these experiments, it was possible to determine the approximate genomic regions contained in each cosmid (Fig. 10). The cosmids hPEA3g.4 and hPEA3g.7 do not overlap exons 2 or 3 of the human genomic PEA3 locus, and sequencing analysis revealed that they contain exon 9 and approximately 40 kb of sequence downstream from this point. The cosmid hPEA3g.1, and its two identical counterparts, hPEA3g.5 and hPEA3g.6 appear only to contain sequence upstream from exon 6. The three other independent cosmid isolates, hPEA3g.2, hPEA3g.3 and hPEA3g.8 must, at a very minimum, overlap exon 3 through exon 6, however their precise endpoints are yet undefined.

c. Subcloning of DNA Fragments From the Human PEA3 Locus

One DNA fragment from hPEA3g.1, found to hybridize to the 5' probe and three fragments from hPEA3g.7, found to hybridize to the 3' probe, were isolated and subcloned into the vector pBluescript KS II. The rationale for subcloning regions of genomic human PEA3 was to aid in the manipulation, sequencing and FISH mapping of the human genomic

Figure 10. Human Cosmid Isolates - A Summary

The exons of the human PEA3 gene are shown in blue at location 17q21.3. Genomic sequences (38-40 kb) contained in each cosmid isolate are shown overlapping the PEA3 locus. The precise endpoints of hPEA3g.2, hPEA3g.3 and hPEA3g.8 are not known, as indicated by hatched bars.


17q21.3
HUMAN PEA3 GENE
10 kb
hPEA3g.4 / hPEA3g.7
38 - 40 kb
hPEA3g.1 / hPEA3g.5 / hPEA3g.6
38 - 40 kb
hPEA3g.2
38 - 40 kb
hPEA3g.3
38 - 40 kb
hPEA3g.8
38 - 40 kb

PEA3 locus. The regions which were successfully subcloned are indicated in Fig. 11 as a blue bar overlapping the genomic region which it contains. The name of each construct appears above each bar.

d. Chromosomal Mapping of PEA3 to Human Chromosomes by FISH.

In order to determine the chromosomal location of PEA3, two subclones of the human PEA3 locus, hPEA3g.RI (1.5) and hPEA3g.RI (1.0) (Fig. 11) were sent to the FISH Mapping Resource Centre at the Hospital for Sick Children, Toronto, to be used as probes. The probes were labeled with biotin and hybridized to normal human lymphocyte chromosomes followed by detection with FITC. The DAPI banded chromosomes and the PEA3 hybridization signal (yellow) at 17q21.3 are shown in Fig. 12a.

To determine if mouse PEA3 lies on the homologous mouse chromosome, a cosmid containing the entire mouse genomic PEA3 locus and 25 kb upstream of the first exon was also mapped by the FISH Centre. It was found that mouse PEA3 is located on mouse chromosome 11 which corresponds to chromosome 17 in humans (Fig. 12b).

e. Sequencing of the Human PEA3 Promoter

The strategy which was used to sequence the human PEA3 locus was three-fold. Initially, in was possible to design primers based on the sequence of the human cDNA. Anticipated splice junctions were deduced by comparing the cDNA sequence with the mouse cDNA of which splice sites are known. Sequencing was performed by the Mobix Central Facility using an ABI 373 automated DNA sequencing apparatus and each primer generated 500 bp of sequence data on average.

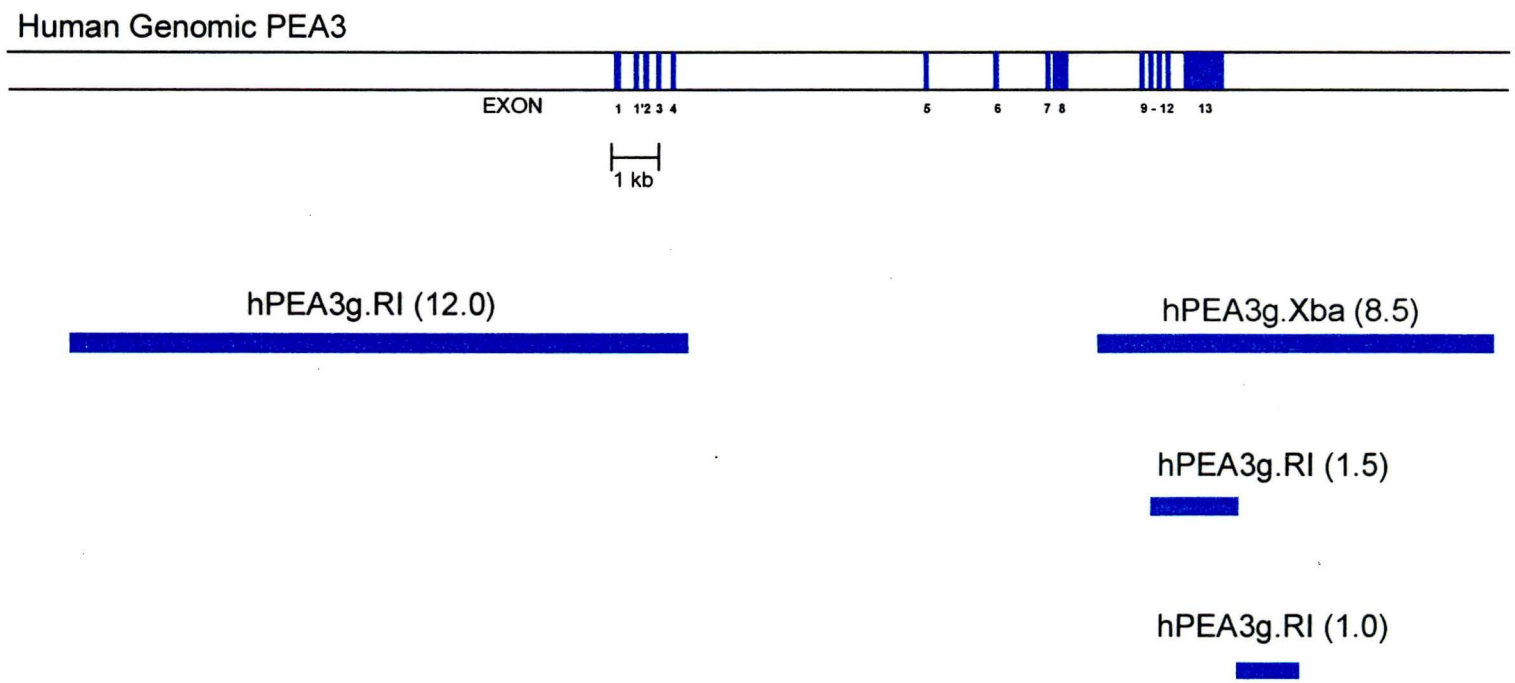

Human Genomic PEA3
EXON
1
1'2 3 4
5
6
7 8
9 - 12
13
1 kb
hPEA3g.RI (12.0)
hPEA3g.Xba (8.5)
hPEA3g.RI (1.5)
hPEA3g.RI (1.0)

Figure 12. FISH Chromosomal Localization Results

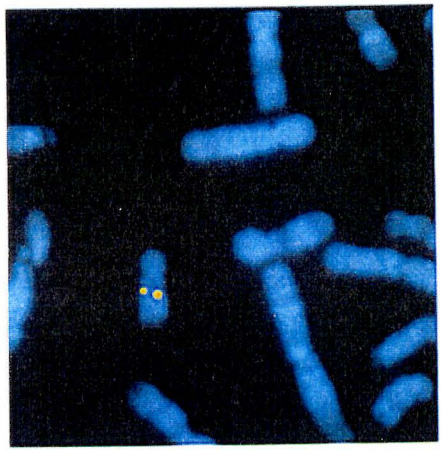
A. Localization of Human PEA3 to 17q21.3

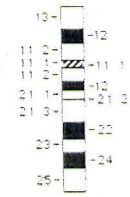
In situ hybridization signal is indicated in yellow and corresponds to chromosome 17 at q21.3. Genes located in this region are listed centromere to telomere.

B. Localization of Mouse PEA3 to 11 d

In situ hybridization signal is indicated in yellow.

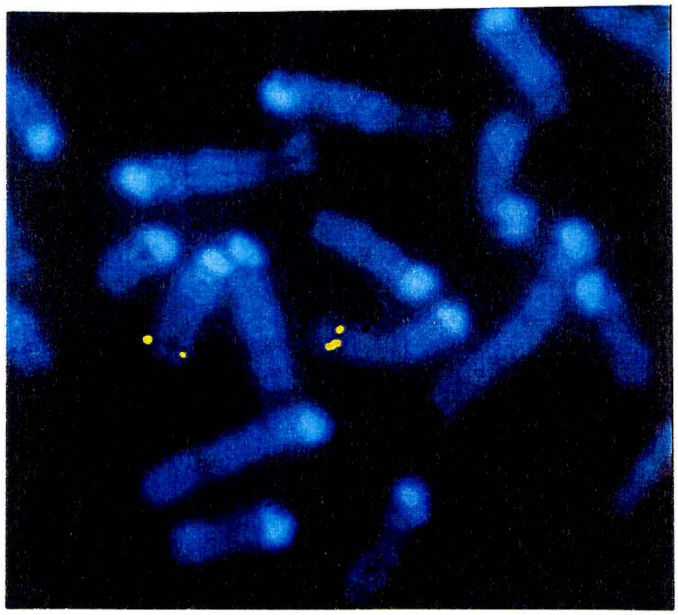
A.





Acyl-CoA desaturase
Proton Pump
a-N-acetylglucosaminidase
EDH17B2
Novel gene
Novel gene
Novel gene
Gamma tubulin
Novel gene
Enhancer of Zeste
HMG17 pseudogene
Novel gene
Novel gene
Ki antigen
Glucose-6-phosphatase
L27 ribosomal protein
Vesicle amine transporter
BRCA1
IAI3B B-box
RNU2
Endogenous retrovirus
VHR phosphatase
Novel gene
MOX2
PEA3
HRH1 helicase

B.



For regions such as the putative promoter for human PEA3 where there does not exist sequence information, a primer walking strategy was used. For sequencing this region, primers were designed starting at the most 5' end of the cDNA and after each successive round of sequencing, another sequencing primer was designed and used for continued sequencing. The third method used to sequence the human gene was by sequencing the subclones by using primers, such as T7 and T3 which are specific for sequences surrounding the multiple cloning sites in various vectors. After numerous sequencing reactions on both strands of the DNA, the data was compiled using the Seq Man subprogram of the DNAStar™ software package. The sequenced regions of the human PEA3 locus contain all exons and three kilobasepairs upstream of the putative promoter region. The sequence of the human PEA3 locus is contained in the Appendix.

f. Computer-Assisted Sequence and Alignment Studies of Mouse and Human PEA3

In order to determine if there were any sequences conserved within the putative promoters of both mouse and human PEA3, three kilobasepairs of their putative promoter regions were aligned using the Meg Align program of the DNAStar™ software package (Fig. 13). Matches in the DNA sequences are indicated by vertical bars. The highest amount of sequence similarity surrounded the transcription start site in exon 1 and proceeded upstream approximately 400 basepairs and at least 700 basepairs downstream from this region. Interestingly, further upstream of this region there exists other stretches of sequence similarity between -520 to -550 and -1540 to -1600. In order to determine whether these stretches of sequence similarity correspond to any known transcription factor

Figure 13. Genomic Alignment of the Putative Promoter Regions of Mouse and Human PEA3

Three kilobasepairs upstream of the first exon of mouse PEA3 was aligned with human genomic PEA3 sequence using a Martinez/Needleman-Wunsch algorithm. Matches are indicated by vertical lines and dashes indicate where gaps have been introduced. Exons in the mouse sequence are underlined.

```
Mouse  GAATTC--TG-GGTCAGACT-GAGTTGAGAGACCCTGTCTCAAAAACATGGTAAGGTA-AGTAGGAAGGAAGGA---AGGAAGGAAGGAAGGAAGGAA--
        | | |  || || || ||  |||| || | || |||||||||||||    ||   | |  |  ||  ||  |   || ||  | | | ||  || |
Human  CACTCCAGTGTGGGCA-ACAAGAGT-GA-A-ACTCTGTCTCAAAAACAA---AA---ACA--A--AA--AA--ACTCAG-AACCA-G-A-GG--GGCACT

Mouse  GG-A-----AGGAAGGAAGGAAGCCACAACAAACCATGAGCTGAGAGACGGTGCATCAGT-C----T-AGAGCACTTGCTG--CT-TT-TTCA--GAGGA
       || |     || | | | ||   || || |   |  | | |||| ||    |||  |||| |    | | || || | |||  || || |||   |
Human  GGCATCCCTAG-A-G-A-GGT--CC-CATC---C--T-A-CTGA-AG----TGCC-CAGTACCAACTTAAAG-ACCTACTGGACTCTTCTTCGTTG----

Mouse  CCTGAGTTCTGTTCCCAGCACCTACATCAGACGGCTCACGACTGCCTGTTACCTGCCGCATCAGGAAATCCAGTGTCTCTGGCCTCTGTAGGTACCCACA

Human  ----------------------------------------------------------------------------------------------------

Mouse  CAAACCTACACAAACACATAAATAAAAATTAAATTAAAAATAAATAACGCAAAATAAGAAAAAAAACGTTATAAAAATTGTGTGGGGGAGAGTCAGCCTC

Human  ----------------------------------------------------------------------------------------------------

Mouse  TTTGATGGCTTCAGAGCCAGCTTGCCAGTTACATTCAGGCTTTCTTTCTTCCAAGCCTGTATCAGGGATCTCCCAGCTCCCAGGGCAAACAGTGGGTTCC

Human  ----------------------------------------------------------------------------------------------------

Mouse  CTTTGTATAAAAATTTACATATCTAAAGTCCACATTTGCATTTTATAATTTAAATTCCTAGGAAACAGGTTAGAGTCCAGCCCCCCAATCTACTGGAGGA

Human  ----------------------------------------------------------------------------------------------------

Mouse  AGAGAATATTTACCTGCCCAGCTTCCCCCAGAGCTGTATTGGAGGAGGCAGGAAGCATCGGAAGTCATAAAGGAGCCCTGAGCATCCTGCATCTGAGGGA

Human  ----------------------------------------------------------------------------------------------------

Mouse  TGCTGGTCAGGTCAGGTTTGGGAGTGTCCTCATGTGCTCCTTAAAGGATCTCTCTTTGGAGTACCCTCAATGGGGGGGGGGTTTCTATGTCACCTCTTGT

Human  ----------------------------------------------------------------------------------------------------

Mouse  AAAATGAGGTGAGGGTGGCAGGCGGCAGGAACTGGTGAAAAGCGTGCACACAAGGCTGGATGCTTCAATGTCACTTAACCCACATCTCCAGCCTCTCTCT

Human  ----------------------------------------------------------------------------------------------------

Mouse  TCCTTCCTCTGCTTAACAACCTTGGACTCCATTGGAGAAGAGGTCCAATCTCCCTTGGAGAAATGTCATTGATCATTGATCATCATCTGAGTGTTCAGTC

Human  ----------------------------------------------------------------------------------------------------

Mouse  CCAGGCCAGGGCAGGTGTGGCAGCTGTAGCTGCAGCTGCAGTTGGTTCTCTTGCTGGGTCTATAGCTTGAGCCAGAGATAGGGAGGTAGGCTTCTTTGGT

Human  ----------------------------------------------------------------------------------------------------

Mouse  CTCTGCCAGACCTGGGAAGCAGAGGAAATATTAACTAGCCAGGGCTCTCCAGAGAAGTGGAGGAACTCCGAGTCAGACAGGCAGCCCCGGACAGGCCTCA

Human  ----------------------------------------------------------------------------------------------------
```

```
Mouse  ACTGGAGCACCCAGTGCTCACTCAGGGACCCACCTGAACCTTCATGGAGACCACATCAGAGAT-CTTTTCCTAG-AACTGCCTGC-CT-TCCTCC-CC-C
                                             ||||||||||||||||| || | | || || || ||  |||  | || |||||| || |
Human  ----------------------------------------TCATGGAGACCACATCAAAGCTGCCTT-CC-AGGAAGAGCC--CTCTGTCCTCCTCCTC

Mouse  -ACTTCTCCTGGTCTAGGAACTCACACCCCTCCCTAAACTGGCTCCATCC-GCCCCAAGCCGGGATTTGCTTGGGCTTGCCTGTCT--TTCCTTGGGCAG
        ||||||| ||| | | ||||||||||   || ||||||||| ||||||| |||| |||| |||||||||||||||||  | |||  ||||||||||||
Human  CACTTCTCTTGG-CCA-GAACTCACACTTGTCTCTAAACTGGTTCCATCCAGCCCAAAGCTGGGATTTGCTTGGGCTTGT-T-TCTCCTTCCTTGGGCAG

Mouse  GGTGCCTGCGGGGCTCCGTTGCTTAGCTTTTTATTGTAGCTGTTCACCTG-ACAGGCATTTTGTGCAACTCT--TGAG--TGCGAAGCGAGGCTCAAGGA
       ||| |||||   ||||  ||  || | | ||||||||| || ||||| |  ||| ||||||  ||| || ||  | ||  ||| || | ||||| | |||
Human  GGTTCCTGCTAAGCTCAATTAATTTGGTCTTTATTGTACCTATTCAC-TACACAAGCATTTCATGCCACCCTCATAAGCATGCTAA-C-AGGCTGA-GGA

Mouse  AGGAGTCCTCCTCTG-GGCCC------C-C-GCAGAGATCCCCATAACT--ATA--TAGGGATTTCTCTAACA-G-GGAAGAGAATGAAAGATA-CCTCA
        ||| |  |  || | ||||       | | | ||||     | |  ||  |||  || |||| ||| || || | || || || ||||| | | || ||
Human  -GGACT--T--TCAGAGGCCAAATGAGCTCTGGAGAG-----C-T--CTCCATAGATAAGGATGTCTATACCATGAGGCAG-GA-TGAAAAACATCCACA

Mouse  CCATATATATGGACT-GACCCAGGCTCTCATGAAAAGGGGGGTCCCTAGAGGAGA--G---C-GGC---CA-G-A-GG-TTCAAACTT--TAA-AG--A-
       ||| | |||| | |  ||   | | | | || |   ||     |  || |||| |  |   | |||   || | | || |||   ||     | ||  |
Human  CCAGAGATATAGTCAAGA---ATGTTGT-AT-AT--GG-----C--TAAAGGA-ACTGATTCAGGCTATCATGGAAGGCTTC---CTGGAGGAGAGGGAC

Mouse  A-----CA-G----TT-TGG--TAGGTA-A--A------G---GGGAGG---AGTT---TGTGTGGCAGAAAGAGCCCAG-T-TCA--GG--TG--CC--
       |     || |    || | |  ||  || | |      |   |||| |   | ||   | | || || || |  |  || | |||  ||  ||  ||
Human  AATAGACATGAGCCTTATAGAATAT-TACAGCACCTATAGAAAGGGAAGCTTACTTCCAT-T-TG-CACAA-G--C--AGATCTCAATGGAATGAGCCCT

Mouse  -AGGGGGT-TAAG--GTTTTTGTC-------------------------------------------------------CCTTCTCTGA--CTCCTACAGG
        | |     |||   |   ||||                                                         ||||||||||  |||| || |
Human  TATGAACACTAAACAGCAATTGTATGTTTTAGCCTTAGTTCTGGCCAGACAGGCCCAAAGAGCTTACATGTGTCATTTCCTTCTCTGAAACTCC-ACTG-

Mouse  CTGGAAGAGATTGCATGCACCTTTTCTTTCTCAGAAGTA-CTG-T-TT-GGTTAGACAGATGAACTTTGA-AAGC-AGC-CCTG-TG-C--GGC-TGTGG
       |            | | ||  |   || |   | ||||  | | | || |||||||    |||| |||| |||| | | |||| || |  ||| | ||
Human  C------------C-T-CA--T---CTGT---A-AAGTGGCGGGTGTTAGGTTAGA----TGAA-TTTGGGAAGCCA-CTCCTGCTGGCCTGGCATTTGA

Mouse  GTC---CTT--GTG---A-C-------GGCATC-TAAAGCAACC---CAGATCC-CAAGGGAGGTAGTCTTT--CTC-ACACTT-GC---TTGGCGCCCA
         |   |||  |||   | |       ||||   ||||  || |   | | ||| ||| |||| | | | |   ||  | || | |    ||| | |  |
Human  A-CATTCTTTTGTGTTTATCACAAACAGGCAGGATAAAA-AATCAGGCTG-TCCTCAAAGGAG-T-G-CATAGGCTGGAGACATAGAGAATTGTCTC--A

Mouse  GCACCA-GTC-T--GACACAGCAGAT--AT-TACTAAG-GC----AGTGATGCTCGA-ATGACAAA-CACA--AGATCAACCTCTTGGACAAAATT----
       |||  | | | |  ||    |  |||  || | ||  | ||    | ||  || | | | | | || |||   || ||  || ||    || |  |
Human  GCAAAAAGGCCTTTGA----G--GATTCATGTCCTTTGTGCCCCCA-TG--GC-C-ACA-GCCCAAGCACCCCAGGTC--CC-CT----CAGATGTCTGT

Mouse  ---GG-------------------------------------------------------------------------------------------
          ||
Human  CTTGGACATAGCTGCTTGCCTGAAACCCAGCACTGTATCTGACACGATAAATGTTTATTGAAAGAAGTGGTGAATGGATGAGATGACAGAGACTATGACC

Mouse  --------------------AACCCATCCTG--TC----T--T-------TT-GA---T----C-----------------A-----------------C
                            |||||||||||  ||    |  |       || |    |    |                 |                 |
Human  AGCTTCTTAGCTAAAACTGAAACCCATCCTGCCTCCAGATCATCAGAACCTTTGGGGCTTTTGCTCCTTTTGCTGGCTCTCAGATAAGAGGGAAGATGGC
```

```
Mouse  ----------C-A-A------G-----TCA--CT-----TG---------GGTTTCTGCTTCCTATCA-TA-A--G-GG-T-A-AGATGGCAGC---G-A
Human  TGAGAGGGAGCNACACTTTTCGGGAAATCATTCTACCCCTGGGGGCTTCTGGTTTCTGCTGCA-ATCAGTAGAAAGTGGGTCACAAAA--CAGCTCTGCA

Mouse  GAGAGACATCCTTGAGCCTCTTGTTGCAAAGCCTCGAATGAATATAAGC-G-G---GTTGA-AT-----GAT-A---AGCGG-GTCA-CAGAATAAGCAA
Human  GAGA-ACA-----GA-CC-CTTGAGGAAAAGCCT-G---GA-T-T---CTGTGCCAGTTTACATTTATTGATCACCTA-CTGTGT-ATCAGT-T--G---

Mouse  CTCTACAGATCATA-GACCCGTG-AGA--A-G-C--C-G-GT--T-T---G-AGCCAGT-TAA---ATTT-A--CTG-AGACC--A-ACTGTGCCTCCAT
Human  -TC-ACATAGCTTAA--CTCATTTA-ATCATGGCTTCAGAGGGGTGTACAGGAG--AGTGTGAGGGATTTGAGGCTGTAAACCCCAGAG-GTGC-TG-AC

Mouse  TTGTTTTCAAGTAG-C-TC-GG-G-GTCTCAGAGGGCTGCA-CAGAAGACCT-TGGTG-GCTTTGAAAGCCTAATGCCCCTCAGTGACAAGCTGAGTTAC
Human  TAGCTG--A-GT-GACCTCAGGAGAGT-T-A-----C--CATC------CCTCTC-TGAG-TTTCAT----T--T-----T---T-A-A---T-A-T--C

Mouse  CTCAGGTGAGGAGGCCATGGTGTCTTTCTTAGCTTTCTTCCATTTTCATTTTGTAAGTAAAATGGAAACCGCTCCGCAGCACCACGTTATGGGAAGGAAA
Human  -T--G-T-A--AG---ATA----CAAACG-A--T--C--C-ATTT-C-----G---------------CCGG-------------GTTATGGGAACGAAA

Mouse  TTAGCACGTGGAGGCAGAGCGCTTCATAAATGCTAGTTATTAT-G--------G---------------T----TC-CT-CTGGGCGCACAAGTCCGTTT
Human  TGAGCACGTGGAAGCGGAGCGCTTCATAAATGCTAGTTATTATTGTTACCTCTGGGCACACAAGCCCGGTCGCCTCCCTCCTGGGCGCACA---C-----

Mouse  GCCTCCCCCTCCCCCCAGGCACACAGACGTGTTCACACTCCCACG-AGAT--AGATGCACAGAGGCGTGCTG-A-ACGCCCGCGCA-AGGAGCTCA-GGT
Human  ----------AG--ACA------TGTTCATACTCCCACGTACACGCAGATGCAGAGA--CGTG-TGCAGACGCCAGC-CGGAGGAGCACGCGG-

Mouse  C-CTTAGACCCAAAACCCAGGTTG-GA-ACCCGTGGAGAAGCTGCCGGGTCTTCAGCTTCNGTCT-TTTTT-TCTCAATTCTTTATGAATGAACTAAAGC
Human  CGCATCGACCCAAGACC--G-TGGCGAGACG-GTGAAGAAGCCGCGGGGTCCTCGGCTTC--TCTCTTTTTCTCTCGGTTCTTTATGAATGAACTCCAGC

Mouse  CGAAAGGAG-TTTATTTTTTTATGAATGGAAGTCCAAAA-AGTGAATGAAGCCAGGAGCCAGCCCCTACTTTCTGATTGGTCTTCGTGGCTGGGAAACTC
Human  CGC--G-AGTTTTATTTTTTTATGAATGGGAG-CCGGAGCACTGAATGAAGCCGGGAGCCAACCCCTACATTCTGATTGGCCCCCGCGGCTGGGAAACTC

Mouse  CTCCCTCAAGCTCAGCTAGCCG-A-G--TCTGAGGATTGGCGGGCGTGCCCGGCCCTG--GGGTTT-CATTCACAAAAGTCAATGAAACAAAGGGAA-C-
Human  CGCCCCCCGGC-CAG---GCCGGACGACTCTCCGGATTGGCGGGCGTGCCCAGCCCTGCTGGGGTTCCATTCACAAAAGTCAATGAAACAAAGGGAAACG

                                                                                                EXON 1
Mouse  -----CGGTGGGGAGGGGCGGGGGAAGAGCGAGAACGGAGGCCAAGGCAAAGGAAATGCACCAATCAGCTGCTCCCCCGGGCTCACAACTGTCTGCTGCG
Human  GGGGGCGGTGTGGCGGGGCGGGGG--GTG-GAGAGCTGAGGCCGAG-CGAAGGAAATGCACCAATCAGCTGCTCCCCCGGGCTCACAACTGTCTGCTGCG

Mouse  CCCGGAAAACAAGCCGGGGCTCTGGGGACCCGGGGCTCAGG-CCGCCTCGCTCCGGCCTAGCCCCGC--CACCTTAGTTGTGT-CATCCCCCGGGCATGC
Human  CCCGAAAAACAAGTCGGTACGNTGGGGACCCGGGGC-CGGGGCCGCCTTATTCCGGCATAGCCCCGCGGC-CCTC-GTTGCGGGC-TCCA--GGGCATGC
```

```
Mouse  I-GAGCATCCCCCCGCGGCTCCGGCACAGACGCCC-GGACCTCAGGTGAAGCT-CGGGAAGGGAA-GGAAGGGAGAGTGTCTTTGGATTTCCCCGCTTCA
       | | | | ||||||||||||||| || ||||||||| || |||||||||| ||| ||||| ||||| || ||||||||||||| || ||| |||||| ||
Human  TCG-GGA-CCCCCCGCGGCTCCAGCCCAGACGCCCCGG-CCTCAGGTGAGGCTGCGGGA-GGGAAAGGGAGGGAGAGTGTCTGTG-ATT-CCCCGCCTCG

Mouse  TCTCCCA--CACACCACACTCTCTCTGTTCGGCCCGCTTCCGCACATTCCGGGGCGGGGGTGGGAGGCGTC-GGGATTCGAGGC-TGGAGGGGGCGCCTG
        |||| |  ||||||||||||| ||||   |   |||| |||||| ||||| |||||||||||||||| | | ||||||| |||  ||||||||||||  |
Human  CCTCCAATGCACACCACACTCACTCTCACCAC-CCGCCTCCGCAGATTCCAGGGCGGGGGTGGGAGGTG-CTGGGATTC-AGGGTTGGAGGGGGCGCGGG

Mouse  AGTTATGGGCGTGCGAATGGTGTGCGC--CGCGC-G-------GCGCGC-C-CACATACACACACACACACATACACACGGG---------ACACACACA
       |||| ||||| |||| ||||||||||  ||| | |       |||||| | | | ||||||||||||||||| ||||||            |||||||||
Human  AGTTGTGGGCATGCG-ATGGTGTGCGCTTCGC-CTGTTCCTCCGCGCGCGCGCGTACACACACACACACACACACACACACACACACACACACACACACA

                                EXON 1'
Mouse  CGCACACACACTTTATGTAACC-AGTT-GGGCTG-AGCAGGGCTGCAGAAAGCAGACACAC--CAACCCCAGCTCAAGGGAGCAGGTGGGACCTC-TTT-
       ||  |       |||||||||| ||   ||| |  |||||||||||||||||||||| | ||  | | ||| ||||  |||||||||||||||||| |||
Human  CGTTC-------TTATGTAACCGAGCCCGGG-TAAAGCAGGGCTGCAGAAAGCAGA-A-ACGGCGAGCCCGGCTCCTGGGAGCAGGTGGGACCTCCTTTG

Mouse  -TCG--GGAGAGGAGTGGCGGTGGAGCGGCTCGGCC-GCGCAGG--------TC-CCA--GC------GAGTCGGAGAAGTTTCGCTGTC---C---GAT
        ||   || | ||||||| |||||  || ||||||| | |||||        || |||  ||      |||||||||||||||||| |||   |   ||
Human  GTCTTTGGCG-GGAGTGGGGGTGG--CG-CTCGGCCTG-GCAGGCGGAAACCTCGCCATCGCCCCAGTGAGTCGGAGAAGTTTCGCCGTCGGACCCAGAG

Mouse  CC-----A-------------------A----AGCC-T----GGN-G-------ATC---GCCCCCTCCCCCCCCGCAGCGCTCCCGGGACCGCGCGCGA
       ||     |                   |    |||| |    ||  |       | |   ||||||||||| ||  || | || |||  || | | ||
Human  CCTTGGGACTGCCCCCTCCCCGCCCCCATCGCAGCCCTCCCTGGACGGTGTGCGAACGCAGCCCCCTCCCCCACC--AG-G-TCTCGG--CC-C-C-CG-

                                EXON 2
Mouse  ACTGCCTCCCCCCTCCCAGGTCTCTGCCTCCGCGCGGGGGCCCGGCCCTGTGGCCGGAGGGAGCGGCCGGATGGAGCGGAGGATGAAAGGCGGATACTTG
        ||   |           ||     ||| ||| ||     |  |    || |||||||||||||||||||||||||||||||||||||||||| |||||||||
Human  -CT---TG----------GG-----GCC-CCG-GC-----C--G----TGCGGCCGGAGGGAGCGGCCGGATGGAGCGGAGGATGAAAGCCGGATACTTG

Mouse  GACCAGCGAGTGCCCTACACCTTCTGCAGC
       ||||||| |||||||||||||||||||||
Human  GACCAGCAAGTGCCCTACACCTTCTGCAGC
```

consensus sequences, the human PEA3 promoter sequences was searched for transcription factor consensus using the MatInspector® program in exactly the same manner as discussed previously for the mouse PEA3 promoter region. Consensus sequences found to be conserved between the mouse and human genomic sequences are shown in Fig. 14.

Figure 14. Consensus Transcription Factor Binding Sites Conserved Between the Putative Promoter Regions of Mouse and Human PEA3

A. Detailed Map of Conserved Consensus Sequences

Putative transcription factor consensus sequences which appear to be conserved between the mouse and human sequences have been identified. The exon sequences have been underlined.

B. Schematic Map of Conserved Consensus Sequences

Putative transcription factor consensus sequences which appear to be conserved between the mouse and human sequence are identified. The exons have been identified by filled boxes.

-1749

CCGGACAGGCCTCAACTGGAGCACCCAGTGCTCACTCAGGGACCCACCTGAACCTTCATGGAGACCACATCAGAGATCTTTTCCTAGAACTGCCTGCCTTCC

-1647

TCCCCCACTTCTCCTGGTCTAGGAACTCACACCCCTCCCTAAACTGGCTCCATCCGCCCCAAGCCGGGATTTGCTTGGGCTTGCCTGTCTTTCCTTGGGCAG

c-Ets-1/PEA3 c-Ets-1 c-Ets-1/PEA3

Ik-2 CP2

-1545

GGTGCCTGCGGGGCTCCGTTGCTTAGCTTTTTATTGTAGCTGTTCACCTGACAGGCATTTTGTGCAACTCTTGAGTGCGAAGCGAGGCTCAAGGAAGGAGTC

Sox5/SRY

-1443

CTCCTCTGGGCCCCCGCAGAGATCCCCATAACTATATAGGGATTTCTCTAACAGGGAAGAGAATGAAAGATACCTCACCATATATATGGACTGACCCAGGCT

-1341

CTCATGAAAAGGGGGGTCCCTAGAGGAGAGCGGCCAGAGGTTCAAACTTTAAAGAACAGTTTGGTAGGTAAAGGGGAGGAGTTTGTGTGGCAGAAAGAGCCC

c-Ets-1

-1226

AGGGGTTAAGGTTTTTGTCCCTTCTCTGACTCCTACAGGCTGGAAGAGATTGCATGCACCTTTTCTTTCTCAGAAGTACTGTTTGGTTAGACAGATGAACTT

-1029

TGAAAGCAGCCTGTGCGGCTGTGGGTCCTTGTGACGGCATCTAAAGCAACCCAGATCCCAAGGGAGGTAGTCTTTCTCACACTTGCTTGGCGCCCAGCACCA

c-Ets-1/PEA3

GTCTGACACAGCAGATATTACTAAGGCAGTGATGCTCGAATGACAAACACAAGATCAACCTCTTGGACAAAATTGGAACCCATCCTGTCTTTTGATCACAAG

-826

TCACTTGGGCTGCTTCCTATCATAAGGGTAAGATGGCAGCGAGAGAGACATCCTTGAGCCTCTTGTTGCAAAGCCTCGAATGAATATAAGCGGGTTGAATGA

c/EBP β

-756

TAAGCGGGTCACAGAATAAGCAACTCTACAGATCATAGACCCGTGAGAAGCCGGTTTGAGCCAGTTAAATTTACTGAGACCAACTGTGCCTCCATTTGTTTT

-656

CAAGTAGCTCGGGGTCTCAGAGGGCTGCACAGAAGACCTTGGTGGCTTTGAAAGCCTAATGCCCCTCAGTGACAAGCTGAGTTACCTCAGGTGAGGAGGCCA

-556

c-Ets-1 c-Ets-1 MyoD/USF

TGGTGTCTTTCTTAGCTTTCTTCCATTTTCATTTTGTAAGTAAGTAAAATGGAAACCGCTCCGCAGCACCACGTTATGGGAAGGAAATTAGCACGTGGAGGC

Ik-2 N-Myc/MycMax

-456

AGAGCGCTTCATAAATGCTAGTTATTATGGTTCCTCTGGGCGCACAAGTCCGTTTGCCTCCCCCTCCCCCCAGGCACACAGACGTGTTCACACTCCCACGAG

Ik-2

-356

c-Ets-1 HSF1 Human PEA3 Footprint

ATAGATTGCACAGAGGCGTGCTGAACGCCCGCGCAAGGAGCTCAGGTCCTTAGACCCAAAACCCAGGTTGGAACCCGTGGAGAAGCTGCCCGGGTCTTCAGCT

Human PEA3 Footprint

-256

Brn-2 Brn-2 c-Ets-1 AP-1

TCCGTCTTTTTTTCTCAATTCTTTATGAATGAACTAAAGCCGAAAGGAGTTTATTTTTTTATGAATGGAAGTCCAAAAAGTGAATGAAGCCAGGAGCCAGCC

HNF-3β NF1

-156

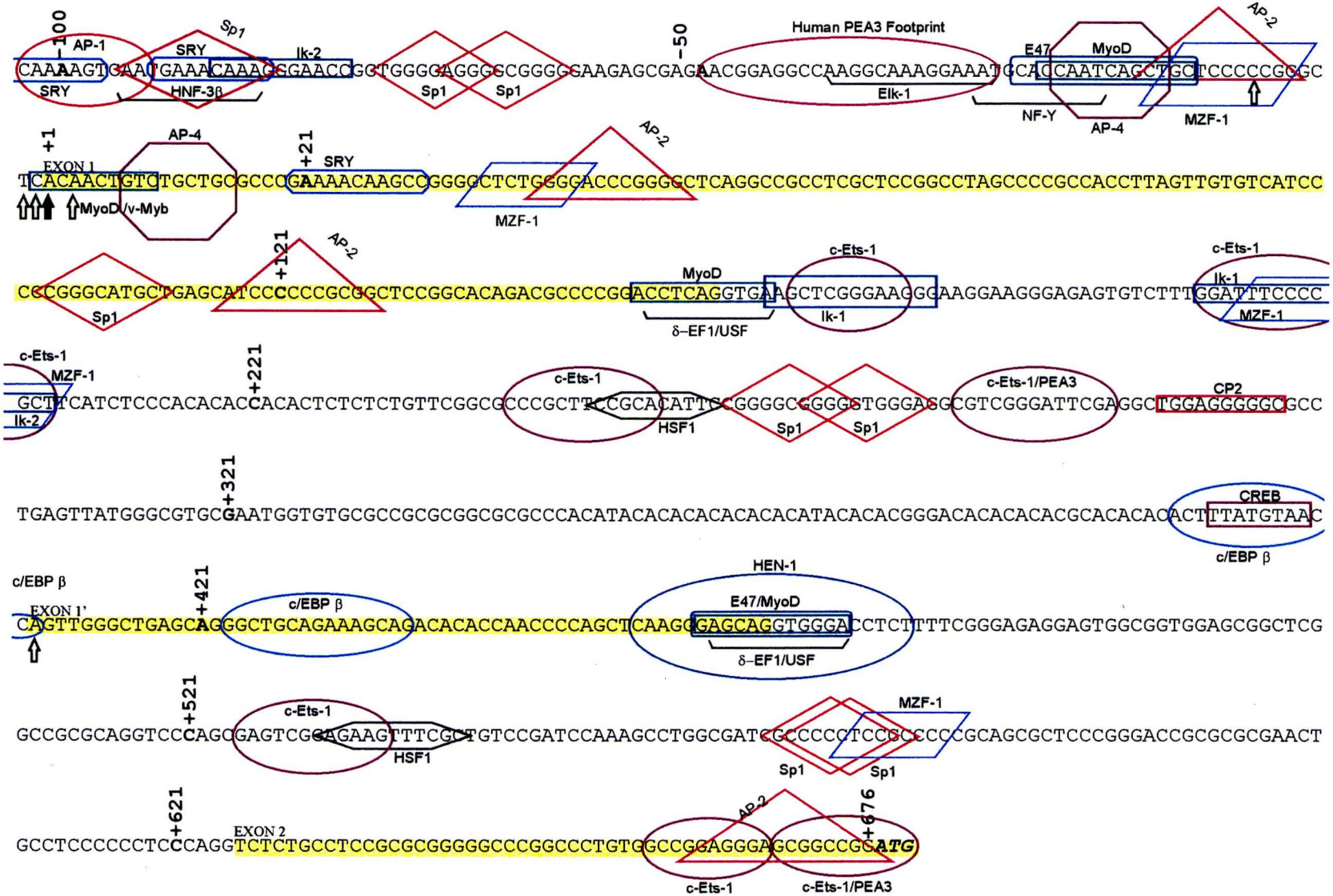
HSF1 c-Ets-1 c-Ets-1/PEA3 c-Ets-1/PEA3 NF-κB

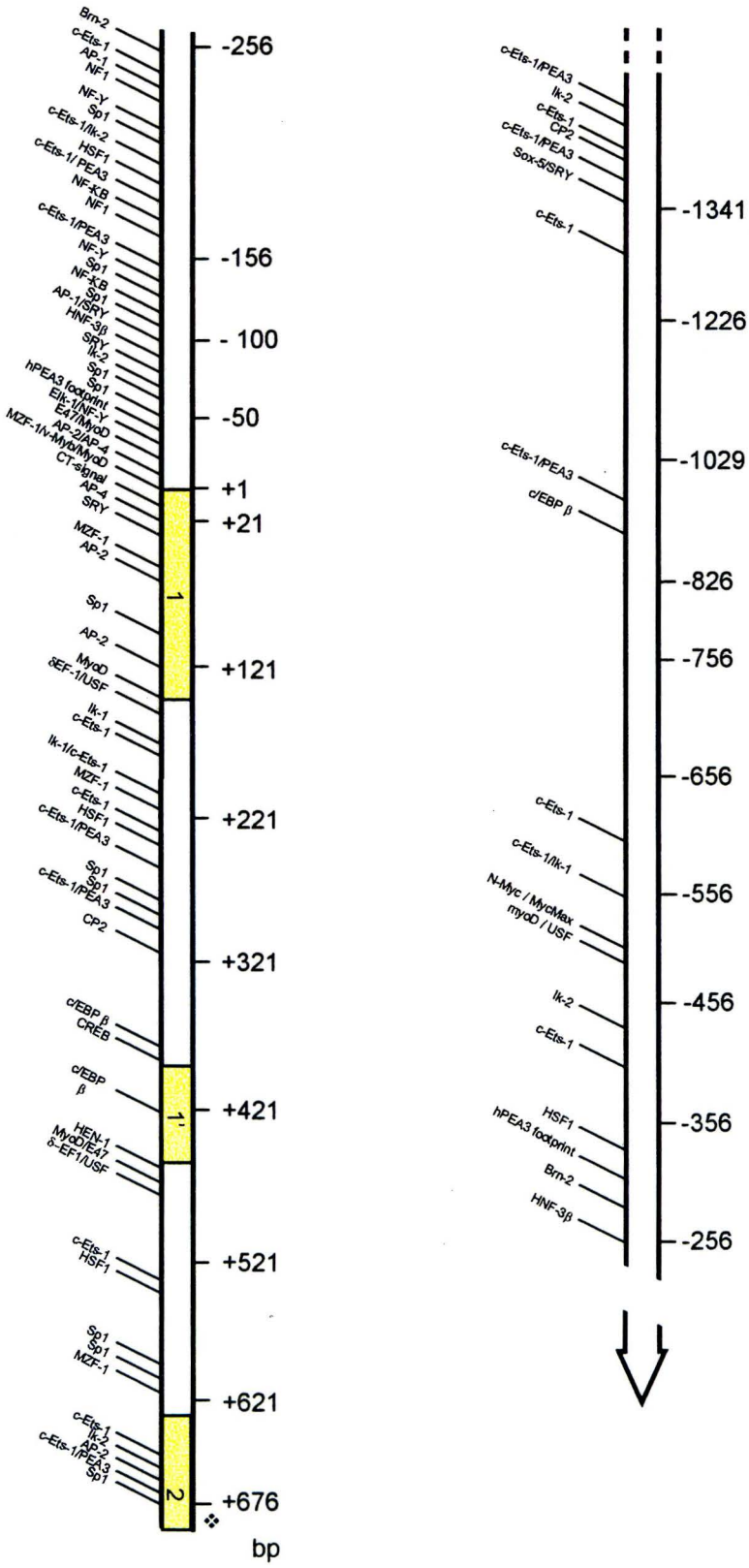
Sp1

CCTACTTTCTGATTGGTCTTCGTGGCTGGGAAACTCCTCCCTCAAGCTCAGCTAGCCGAGTCTGAGGATTGGCGGGCGTGCCCGGCCCTGGGGTTTCATTCA

NF-Y Ik-2 NF-Y Sp1 SRY

NF1/NF-κB


-100
AP-1
SRY
Sp1
Ik-2
SRY
HNF-3β
Sp1
Sp1
-50
Human PEA3 Footprint
Elk-1
E47
MyoD
AP-2
NF-Y
AP-4
MZF-1
+1
AP-4
EXON 1
+21
SRY
MyoD/v-Myb
MZF-1
AP-2
+121
AP-2
Sp1
MyoD
c-Ets-1
δ-EF1/USF
Ik-1
c-Ets-1
Ik-1
MZF-1
c-Ets-1
MZF-1
Ik-2
+221
c-Ets-1
HSF1
Sp1
Sp1
c-Ets-1/PEA3
CP2
+321
CREB
c/EBP β
c/EBP β
EXON 1'
+421
c/EBP β
HEN-1
E47/MyoD
δ-EF1/USF
+521
c-Ets-1
HSF1
MZF-1
Sp1
Sp1
+621
EXON 2
AP-2
+676
c-Ets-1
c-Ets-1/PEA3


Brn-2
c-Ets-1
AP-1
NF1
NF-Y
Sp1
c-Ets-1/Ik-2
HSF1
c-Ets-1/ PEA3
NF-κB
NF1
c-Ets-1/PEA3
NF-Y
Sp1
NF-κB
Sp1
AP-1/SRY
HNF-3β
SRY
Ik-2
Sp1
Sp1
hPEA3 footprint
Elk-1/NF-Y
E47/MyoD
AP-2/AP-4
MZF-1/v-Myb/MyoD
CT-signal
AP-4
SRY
MZF-1
AP-2
Sp1
AP-2
MyoD
δEF-1/USF
Ik-1
c-Ets-1
Ik-1/c-Ets-1
MZF-1
c-Ets-1
HSF1
c-Ets-1/PEA3
Sp1
Sp1
c-Ets-1/PEA3
CP2
c/EBP β
CREB
c/EBP
β
HEN-1
MyoD/E47
δ-EF1/USF
c-Ets-1
HSF1
Sp1
Sp1
MZF-1
c-Ets-1
Ik-2
AP-2
c-Ets-1/PEA3
Sp1
-256
-156
- 100
-50
+1
+21
+121
+221
+321
+421
+521
+621
+676
1
1'
2
bp
c-Ets-1/PEA3
Ik-2
c-Ets-1
CP2
c-Ets-1/PEA3
Sox-5/SRY
c-Ets-1
c-Ets-1/PEA3
c/EBP β
c-Ets-1
c-Ets-1/Ik-1
N-Myc / MycMax
myoD / USF
Ik-2
c-Ets-1
HSF1
hPEA3 footprint
Brn-2
HNF-3β
-1341
-1226
-1029
-826
-756
-656
-556
-456
-356
-256

DISCUSSION

The objectives of this project were successfully completed. This section will discuss the sequences of the PEA3 promoter which govern expression in the FM3A cell line. To support data the data obtained from the luciferase assays, the human PEA3 gene upstream region was sequenced and analyzed for factor binding sites. By compiling the conserved factor binding sites from both species, inferences could be made regarding the role of each in PEA3 expression.

1. Cloning and Detection in the pGL3 Vector System

Although some putative promoter fragments for the mouse PEA3 gene were isolated from the genomic sequence using restriction endonucleases, other methods to isolate sequences had to be used when suitable restriction sites did not exist. PCR was invaluable for the isolation of defined sequences. Cloning into the pGL3 vector was simplified by engineering restriction endonuclease sites into the ends of the amplified DNA fragments. It was necessary to analyze each putative PEA3 promoter clone by restriction endonuclease digestion and sequencing, because the PCR reaction occasionally inserts a mutation into the amplified sequence. The PEA3 promoter-luciferase constructs generated from sequences cleaved from the genomic sequence were only sequenced at the cloning site junctions.

2. Luciferase Assays

The luciferase assay allowed for analysis of the PEA3 promoter region. The luciferase assay was a simple, sensitive and non-isotopic method for assaying reporter gene activity. The pGL3 vector has advantages over other luciferase reporter vectors. The luciferase gene has been modified to contain a Kozak consensus sequence for increased translational efficiency (Kozak, 1991). Peroxisomal targeting sequences have also been removed, thus luciferase is localized in the cytoplasm. The vector backbone contains a poly(A) sequence upstream of the cloning site to prevent any spurious transcription and a SV40 late poly(A) sequence for increased mRNA stability.

The variability in reporter gene activity due to transfection efficiency or cell extract preparation was normalized using an internal control. The plasmid, pSV-βgal, expressed β-galactosidase from the SV40 promoter and enhancer. Usually β-galactosidase activity is assayed by monitoring the hydrolysis of ONPG (*o*-nitrophenyl-β-D-galactopyranoside) by a simple colorimetric assay. This assay is not very sensitive as it only has a detection limit of approximately 1 ng of β-galactosidase (MacGregor *et al.* 1990). By using the chemiluminescent β-galactosidase substrate, AMPGD, 10 fg of β-galactosidease can be measured accurately (Jain *et al.*, 1991).

The Dual Light™ luminescence reporter gene assay system combined both the luciferase and β-galactosidase assays in a single extract. This assay enabled greater precision when normalizing for transfection efficiency since separate assays were avoided. Light from each enzymatic assay was measured sequentially using a luminometer.

The equation for calculating the luciferase activity of each extract was corrected for the number of moles of each plasmid, or luciferase gene copy number, in each transfection because the size of each PEA3 promoter/luciferase chimera ranged from 5 kb to 12 kb. To determine if a statistically significant difference existed between the data obtained between two different constructs, an ANOVA (analysis of variance) test was performed. This method of analysis was very useful in determining the statistical significance of the data obtained from the assay.

a. General Analysis of Promoter Activity

To identify the sequences required for the transcription of a gene, an excellent approach is to clone nested deletions of the putative functional elements upstream of a reporter gene and assaying for changes in biological activity. The first set of unidirectional promoter deletions were made from the 5' end of the PEA3 gene. Following the reporter gene assays on this series, the sequences downstream of the transcriptional start were analyzed for their requirement in PEA3 expression by making 3' deletions. There are three series of 3' deletions based on three 5' endpoints, -356, -156 and the -3 series. It became evident that the 3' deletion constructs must be analyzed with caution. The 3' endpoints of each putative PEA3 promoter region exist following the removal of splice acceptors and splice donors which may lead to aberrant splicing and translational initiation.

In parallel experiments, the human PEA3 sequence was analyzed for evolutionarily conserved sequences in order to help identify functional promoter elements. It is

incredibly important to realize that this analysis will only help suggest which factors might be required. Experiments will have to be designed in order to prove the requirement of these sequences and factors and will be discussed in the following sections.

b. 5' Deletions of the Putative Promoter

Since the activity of the PEA3 promoter increased significantly when the region from -7 kb to -1341 was deleted, this implies that a silencer element had been removed. Although there are no matches to a silencer element, alignment of the mouse and human genes revealed that there were consensus factor binding sites (all matches above 85%) to 3 c-Ets-1, 2 PEA3, 1 Ik-2, 1 CP2 and 1 SRY site. These conserved sequences were located approximately 1600 bp upstream from the start site of transcription (Fig. 14).

The next deletion construct, -1226+676 *luc*, had an activity that was not statistically different from the promoterless construct. As subsequent deletions were made from 1029 to 656 bp upstream from the start site of transcription, promoter activity generally increased as sequences were removed. In this region of the promoter, conserved consensus binding sequences were identifiable. There appears to be a c-ets-1/PEA3 site conserved between -1029 and -826. There didn't appear to be sequences which aligned between mouse and human between -826 and -656, and therefore no consensus binding sequences were mapped. It is possible that factor binding sites only existing in the mouse promoter in this region might account for this slight increase in promoter activity. When constructs having -656, -556 and -456 as their 5' most sequence were assayed there was no significant difference in their activities which means that the sequences between -656

and -456 did not appear to be contributing to the relative activity of the promoter constructs. Analysis of the DNA sequence in this region revealed that there were 1 c-ets-1 site between -656 and -556. Between 556 bp and 456 upstream of the transcription start site there were sequences that were similar in both mouse and human and they correspond to 1 c-Ets-1, 1 Ik-2, 1 MyoD, 1 Myc and 1 MycMax consensus binding site. Deletions of these sequences, however, did not affect transcription of the reporter gene. Since the activity increased significantly when the sequences between -456 and -356 were deleted, it implies that a negative element was removed from the putative promoter that exists between 456 and 356 bp upstream. There were sequences in both the mouse and human promoter which were conserved in this region and they correspond to 1 Ik-2 and 1 c-ets-1 site. There was no statistically significant change in activity when the sequence between -356 and -256 was deleted. Interestingly, this region contained a sequence which was found to bind to the human PEA3 protein (Benz *et al.*, *Oncogene,* in press). Furthermore, it contained a consensus binding sequence for the HSF1 and Brn-2 factors. When the sequence between -256 and -156 was removed there was an increase in reporter gene activity. This implies that there must be a negative element located between 256 and 156 upstream of the transcription start site. In this putative promoter region there are many sequences which are very similar between the mouse and human species. These conserved regions correspond to the following putative binding sites for the following factors: 1 HNF-3β, 1 Brn-2, 3 c-Ets-1, 1 PEA3 site, 1 AP-1, 1 Sp1, 1 NF-Y, 1 NF1, 1 NF-κB, 1 Ik-2 and 1 HSF-1 site.

The luciferase construct with the greatest reporter gene activity in this series of assays was -156+676 *luc*. The mouse promoter region contained in this construct had a very high sequence similarity to human and, in fact, the identity between these two sequences was approximately 90%. Deletions from the 5' end to -100, -50 and -3 only slightly reduced luciferase activity. The sequence elements which were conserved between mouse and human in this region are 9 Sp1 sites, 8 c-Ets-1 sites, 3 PEA3 sites and one PEA3 footprint, 4 AP-2 sites, 4 MZF-1, 4 MyoD, 3 SRY, 2 Ik-1, 1 Ik-2, 2 E47, 2 NF-Y, 2 c/EBPβ, 2 AP-4, 2 δ-EF1/USF, and one of each of the following: HSF1, CP2, HEN-1, NF-κB, v-myb, AP-1, HNF-3β, and Elk-1. In order to narrow in on the sequences which were involved in promoting transcription, the obvious next step was to construct deletions from the 3' end of these sequences and look for changes in activity.

c. 3' Deletions of the Putative Promoter

The series of 5' deletions were assayed separately from the series of 3' deletions. This may explain why the relative strengths of the promoter constructs which were included in both analyses have assay values which vary slightly. For instance, the absolute rank order of -156+676 *luc* and -3+676 *luc* reversed. However, there is still an agreement in that both of these fragments direct a similar level of luciferase expression.

There were three sets of 3' deletions with different 5' endpoints. The 5' endpoints were located 356 bp, 156 bp and 3 bp upstream from the major start site of transcription. For each set of deletions, the longest construct, which is the one which had a 3' endpoint at +676 had the highest activity. It is important to note that this construct

contains two putative promoter regions, one for transcription initiation at exon 1 and one for initiation at exon 1'. In this series of assays, the construct which contains the sequence from -3 to +676 had the strongest activity. The simplest way to evaluate the effects of each deletion from the 3' end was by analyzing the effect of the deletion on each set (-356,-156 and -3) at the same time.

The first observation was that activity was reduced by a factor of two when 55 bp was deleted from the 3' end of the -356 and the -3 sets of deletions. The consensus transcription factor binding sites which were conserved between mouse and human in this region were 2 c-Ets-1, 1 PEA3 and 1 AP-2 site. It is possible that this decrease in activity was either due to the removal of these sequences or the effect of the deletion of a splice acceptor at the beginning of exon 2. Analysis of cDNA clones shows that exon 1 and exon 1' splice only to exon 2 or exon 3 without exception (Smillie, 1993). Since the splice acceptor at the beginning of exon 2 has been removed from these transcripts it is possible that a section of the luciferase gene was acting as a splice acceptor and therefore the reporter gene was not being expressed properly. Thus, the luciferase activity of these constructs, that is, those whose 3' endpoints are at +621, may not necessarily assist in identifying regulatory sequences for directing expression of PEA3. The -156+621 *luc* construct does not agree with the above hypothesis. The 55 bp deletion from the 3' end only slightly reduced luciferase activity. The only explanation for this result is that these 55 bp did not contribute to reporter gene expression in the construct with a 5' endpoint at -156.

For all of the deletion constructs with 3' endpoints at +521 and +421, activity compared to the +676 construct was reduced two-fold for the -356 set, three-fold for the -156 set, and six-fold for the -3 set. The conserved consensus sequences deleted from each promoter between +621 and +521 were 1 c-Ets-1, 1 HSF1, 2 Sp1s and 1 MZF-1. The deletion of these sequences may account for the decrease in activity of all of these constructs. In all of these constructs, putative initiator elements surrounding the start sites of transcription at exon 1 and 1' are present. The deletion of conserved consensus binding sites for c/EBPβ, HEN-1, E47/MyoD and USF/δ-EF1 between +521 and +421 did not have a significant effect on promoter activity. All of these luciferase constructs had relatively low activities. As discussed above, this may be due to aberrant splicing of the mRNA following transcription which disrupted the expression of the luciferase gene. Another explanation could be due to aberrant translational initiation. Exactly what this means is that there are three potential start codons in the intron between exon 1 and 1' and this intron is normally spliced out when a splice acceptor at exon 2 is present. Since the splice acceptor in exon 2 has been deleted, this intron containing the 3 start codons will probably be present in the mRNA. Furthermore, not one of these start codons are in frame with the luciferase gene reading frame. Although the luciferase gene has a perfect Kozak consensus, one of the other start codons has only one base pair mismatch to the Kozak consensus so translation is most probably initiating erroneously. This suggests again that assaying the activity of these 3' deletion constructs may not necessarily assist in identifying regulatory sequences directing expression of PEA3.

When 200 bp deletions are made such that the 3' endpoints are at +221, there was an increase in activity in all three sets compared to the clones with endpoints at +421. The sequence between +221 and +421 contained the following consensus binding sites which were conserved between mouse and human: 2 c-Ets-1 sites, 1 PEA3 site, 1 CREB, 1 c/EBPβ, two Sp1, 1 CP2 and 1 HSF1. Most interestingly, this region contains a CA-repeat (n>8) which was conserved in an intron in both the mouse and human promoter sequences. This CA-repeat, when n>4, has been shown to function as a silencer element (Wu *et al.*, 1994) and seems to function as a silencer element in the PEA3 promoter.

When you remove the 100 bp from the 3' end to generate the three clones with endpoints at +121, activity decreased in all instances. In the case of the series with the upstream endpoint at -3, the activity was actually reduced to an activity which did not differ significantly from the promoterless construct. The sequences, conserved between mouse and human which appear important for transcription, since activity decreased when they were removed are: 1 MZF-1, 1 Ik-2, 2 c-Ets-1, 1 Ik-1, 1 Myo D, 1 δEF1/USF and 1 AP-2. Whether or not these sequences are truly involved in PEA3 expression can not be ascertained based on theoretical binding sites, especially since these sites are based on an 85% match to the consensus binding site. Moreover, not all of the factors may be present in the cells which express PEA3 nor the FM3A cell line in which these assays were performed.

The next set of 3' deletions gave rise to three constructs with downstream endpoints at +21 bp. This deletion restored the activity of the -356+21 *luc* construct to

the activity of the full length (-356+676 *luc*) construct. There was no significant difference in promoter efficiency when these 100 bp were deleted from -156+121 *luc*. The construct, -3+21 *luc*, had an activity which was not significantly different from the promoterless construct. Since the results of this deletion gave inconsistent results for each construct the effects of the DNA sequences between +21 and +121 can not be established.

Two constructs were analyzed which only contained sequence upstream from the major initiation site in exon 1. These two constructs have 3' endpoints which disrupt the Inr element encompassing the transcription start site. Both of these constructs had relative activities which did not differ from the activity of the promoterless construct, PL *luc*. This would imply that sequences upstream from the Inr element have no effect on transcription if the Inr element has been disrupted. If one compares the activity of -356 *luc* with -356+21 *luc* and -156 *luc* with -156+21 *luc* the effect of restoring the Inr element and the addition of the 21 bp was sufficient to increase the promoter activity 5 - 7 fold. The putative factor binding sites contained in the region from +1 to +21 are: 1 Sp1, 1 AP-2, 1 MZF-1 and 1 SRY.

These experiments have demonstrated that sequences surrounding the putative promoter for PEA3 can drive expression of a reporter gene. The following section will discuss the likelihood of the putative factor binding sites being truly involved in PEA3 expression and those experiments to verify these results. The reasons as to why the deletion of highly conserved sequences, such as those between -156 and -3, did not have a significant deleterious effect on transcription in the FM3A cell line is unclear. This suggests that other conserved sequences may also have similar trivial effects.

d. The Effect of the SV40 Enhancer on the PEA3 Promoter

The rationale for cloning the SV40 enhancer far upstream of these putative promoters was to determine whether *bona fide* promoter elements were present, because enhancers will only enhance transcription from real promoters.

Five of the putative PEA3 promoter fragments were selected at random from the collection of PEA3 promoter constructs. The SV40 enhancer sequence was isolated from the pGL3-enhancer vector (Promega) and cloned into the pGL3-PEA3 chimeras. These constructs were then transfected into FM3A cells and their activity then compared to their non-enhanced counterpart. The luciferase constructs, -356+221 *luc* and -156+121 *luc*, were activated 10.7 and 11.3 fold. This activation was quite significant because when you compare it with the fold activation that the SV40 enhancer had on the SV40 promoter in FM3A cells, this was 6.6 fold. This result confirms the presence of functional promoter elements in the 5' end of the PEA3 gene. The constructs, -356+676 *luc*, -156+676 *luc* and -156+221 *luc*, were three constructs which had relatively high levels of luciferase activity in the absence of the enhancer sequence. In conjunction with the SV40 enhancer, their activity increased between 3.6 and 3.8 fold. One explanation for this might be the presence of the $(CA)_n$ sequence between +221 and +421 which is a known silencer element (Wu *et al.*, 1994). As to whether this enhancer activation is genuine, this is unclear. The supposedly promoterless construct (PL *luc*) was activated 7.4 fold when the enhancer element was present. This result seems feasible because the luciferase units obtained for PL *luc* were approximately half of the luciferase activity of a construct containing the TATA element from the Ad2 gene in the same cell line (data not shown).

In other words, PL *luc*, had some promoter activity in FM3A cells. This may be due, in part, to the fact that the region upstream of the luciferase gene in the pGL3 vector possesses quite a few putative PEA3 motifs, and the PEA3 protein is quite abundant in FM3A cells.

e. Factors Regulating PEA3 Expression

To simplify the assessment of the factors involved in PEA3 transcription, only the factor binding sites conserved between mouse and human in the strongest construct will be considered. It should be pointed out that these putative factor binding sites may or may not have played a role in PEA3 expression. Furthermore, there were many more sequences which were conserved between mouse and human that did not have an 85% match to a transcription factor consensus site. It is also possible that the conserved sequences were not involved in regulating the expression.

The luciferase reporter plasmid containing the promoter region from -3 to +676 had the highest relative activity. The sequence elements that are conserved between the mouse and human genes in this region are: 6 Sp1 sites, 8 c-Ets-1 sites, 3 PEA3 sites, 3 AP-2 sites, 3 MZF-1, 2 MyoD, 2 Ik-1, 2 c/EBPβ, 2 δ-EF1/USF, 2 HSF1 and one of each of the following: AP-4, Ik-2, SRY, CP2, HEN-1, CREB, and E47.

The Sp1 sites may have a significant effect on promoting transcription of the PEA3 gene. It has been reported that the binding of Sp1 factors greatly stimulates transcription from the Inr of several genes (Smale and Baltimore, 1989; Pugh and Tjian, 1991). This activation may be mediated by a direct interaction between Sp1 and TAF110 which

recruits TFIID to the template. To assess the potential role Sp1 has on transcription of PEA3, footprinting or electrophoretic mobility-shift assays (EMSAs) should be performed with the PEA3 promoter to see whether these are *bona fide* Sp1 binding sites.

Of special interest, are the putative binding sites for PEA3. Studies on the promoter regions of other ets family members, ets-1 and PU.1, have shown that the presence of their binding motifs have given rise to autoregulatory loops (Oka et al, 1991; Kistler *et al.*, 1995; Chen *et al.*, 1995). PEA3 also transactivates its own promoter in Cos cells (Benz *et al.*, *Oncogene*, in press). The luciferase construct containing the region from -156 to +676 is activated in a dose dependent manner when co-transfected with a PEA3 effector plasmid. Furthermore, PEA3 binds to the mouse PEA3 promoter region and induces a DNase-I hypersensitive site between -49 and -27 and between -302 and -280 (Benz *et al.*, *Oncogene*, in press). These DNase-I footprinting studies were performed on the promoter sequence from -356 to +1. It would now be of interest to assess whether the 3 PEA3 motifs downstream from the transcription start site also bind PEA3. These results help to confirm that a *bona fide* promoter has been isolated for mouse PEA3.

The presence of 8 consensus c-ets-1 sites also suggests that other ets family members might have a role in the expression of PEA3. For instance, there are a number of Ets transcription factors are expressed in the mammary gland including ESX, ERM, ER81, Ets-2, GAPBα, Elk-1 and Elf-1 (Chang *et al.*, 1997, Baert *et al.*, 1997; Scott *et al.*, 1994), which may account for expression.

The presence of 3 AP-2 consensus sites in the PEA3 promoter is also of interest. The AP-2 transcription factor is elevated in all cell lines overexpressing c-*erb*B-2 and c-*erb*B-3 and is presumed to mediate this overexpresssion (Kraus *et al.,* 1987; Skinner & Hurst, 1993). Since PEA3 is overexpressed in 93% of c-*erb*B-2 overexpressing human breast tumor samples, the role of the AP-2 factor in regulating PEA3 expression is of great interest. There are three members of the AP-2 family of transcription factors, AP-2α, AP-2β and AP-2γ (Bosher *et al.,*1996). Preliminary studies have shown that the AP-2 family of transcription factors were able to transactivate the PEA3 promoter (-156+676 *luc*) in the HepG2 cell line (data not shown). The HepG2 cell line was chosen for this analysis since it expresses low levels of the AP-2 family members. Further studies to evaluate the role of AP-2 in PEA3 expression will be to perform DNase footprinting studies or EMSAs with the PEA3 promoter sequence.

The MEF1 factor is involved in the transcription of muscle-specific genes and is formed from the heterodimerization of MyoD and E47 (Weintraub *et al.,* 1989; Lassar *et al.,* 1991). PEA3 has been detected and accumulates in activated satellite cells. Since MyoD accumulation is an indicator of satellite cell activation (Grounds *et al.,* 1992) it is reasonable to ask whether the expression of PEA3 is regulated by the MyoD factor. PEA3 is also expressed to high levels in the C2C12 mouse myoblast cell line (Taylor *et al.,* personal communication). When these cells are induced to differentiate the levels of PEA3 mRNA and protein decreases. It is therefore reasonable to look at whether the expression in this cell line is controlled at the level of transcription. Preliminary promoter

studies in the C2C12 myoblast cell line has confirmed the effect of a silencer element in the upstream region of the putative promoter (Peterson, personal communication).

There are 3 MZF-1 (myeloid zinc finger protein-1) and one CP2 consensus factor binding site conserved between the mouse and human PEA3 promoter in the region between -3 and +676. MZF-1 is thought to be involved in the regulation of hematopoietic development (Perrotti *et al.*, 1997) and CP2, a distant relative of Elf-1 is expressed primarily in hematopoietic cells (Kim *et al.*, 1990). Since PEA3 mRNA is detected in hematopoietic cells (Romana-Spica *et al.*, 1994), these factors may be required for transcription of the PEA3 gene in these cells.

Ikaros genes 1 and 2 (Ik-1 and Ik-2) arise from alternate splicing of the Ikaros mRNA and are regulators of lymphocyte commitment and development (Molnar *et al.*, 1994). PEA3 is preferentially expressed in specific hematopoietic lineages (ie. myeloid, erthroid) but is barely detectable in lymphoid progenitors (Romana-Spica, 1994). Therefore, these factors are probably not important for PEA3 expression.

The following factors are expressed ubiquitously, AP-4, c/EBPβ, CREB, USF and HSF1. The c/EPBβ (CCAAT enhancer-binding protein) recognizes the CCAAT box (Alam *et al.*, 1992). CREB (cAMP response element-binding protein) has been shown to interact with a component of the TFIID complex (Ferreri *et al.*, 1994). USF (upstream stimulatory factor) may be involved with Inr-element mediated transcription although the exact mechanism is not clear (Du *et al.* 1993). And HSF-1 (Heat shock factor-1) is induced to bind following heat shock (Rabindran *et al.*, 1991).

δEF-1 is a chicken transcription factor with lens-specific activity (Funahashi *et al.*, 1991) but is found in other tissues. Perhaps the mouse homologue is involved in the transcription of mouse PEA3.

HEN-1 is a neuron-specific factor that plays a role is development of the mammalian nervous system (Brown and Baer, 1994). This factor may be important for PEA3 expression since PEA3 is expressed in motor neurons in the developing mouse embryo (Laing, personal communication).

The SRY factor is a chromosome Y linked gene and is involved in the development of testis (Nasrin *et al.*, 1991). Perhaps this factor could also regulate PEA3 expression in the epididymis.

The PEA3 promoter was shown to be cell-type-specific (Benz *et al.*, in press). The PEA3 promoter-luciferase construct containing the sequences from -156 to +676 was found to have a very low activity in the COS cell line. This cell line does not express detectable levels of PEA3 protein or mRNA. It might also be useful to evaluate PEA3 promoter activity in a non-oncogenic cell line in which PEA3 is normally expressed, such as an epididymal cell line. A future experiment might be to perform footprinting experiments or EMSAs in cell lines which do express PEA3 and compare these with the COS cell line which does not. This should reveal information as to which sequences may be involved in transcription of the gene. Following along with these studies, mutational analysis could be performed to verify these results. Mutation of the consensus Inr element may abolish transcription initiation at both exon 1 and 1'. In fact, in order to specifically

address the activity of the promoter for transcription initiation at exon 1', deletions should be made at or near the region immediately upstream from this alternate start site.

Future experiments might also be directed at mapping the retinoic acid responsive element (RARE) or the Neu responsive element in the PEA3 promoter. PEA3 expression is down regulated following retinoic acid induction of cell differentiation (Xin *et al.*, 1992). To this end, P19 cells would be transfected with the PEA3 promoter-luciferase constructs and the effect of retinoic acid on deletions of the promoter would be used to map the RARE. In similar experiments the ability of constitutively activated Neu to activate deletions of the PEA3 promoter may be used to characterize the Neu responsive element.

3. Isolation and Sequencing of the Human PEA3 locus

The full length genomic clone of human PEA3 with at least 15 kb of flanking sequence upstream and downstream of the gene was isolated from a cosmid library. The molecular organization of the exons and introns were found to be almost identical to mouse PEA3. Numerous attempts to sequence the entire large intron between exons 4 and 5 failed. This intron is predicted to be 6 kb based on the structure of the mouse PEA3 gene and approximately 3 kb of the central portion of the intron remains to be sequenced. The inablility to sequence this region may have been due to secondary structures present in this region of the gene. Sequences were found to be conserved between the two genes in a region within 1800 bp upstream of the PEA3 gene. A most interesting result was the high level of sequence identity between the two genes from a region 300 bp upstream of

the transcription start site to the end of exon 1'. This region was 90% identical between the mouse and human genes. There were also strings of conserved sequences in the second intron upstream of exon 2. This phenomenon exists in other genes, including the ets gene, PU.1,of which the 500 bp region surrounding the transcription start site is almost 90% identical between mouse and human (Chen *et al.*, 1995). As an additional tool to map critical elements for PEA3 expression, the putative promoter along with the PEA3 gene is being isolated from chicken and pufferfish genomic libraries. The pufferfish genome is an excellent model system since its genome is condensed, thus, the majority "junk" DNA has been eliminated (Brenner *et al.*, 1993; Koop *et al.*,1996). Hopefully, through direct sequence comparison sequences governing expression of PEA3 will be conserved and more importantly, recognizable.

4. Chromosomal Localization of PEA3

Two subclones from the human PEA3 genomic region were used as probes to map the PEA3 gene. At the same time as the FISH analysis was underway at the Hospital for Sick Children, another lab mapped PEA3 to location 17q21 (Isobe *et al.*, 1995). Their result differs slightly from the results presented here with PEA3 mapping to 17q21.3. The reason these results may be different is due to the fact that two different techniques were used for this analysis. Isobe *et al.* used somatic cell hybrid analysis with 26 rodent-human hybrids as a technique to map human PEA3. FISH analysis also mapped mouse PEA3 to the syntenic mouse chromosome 11. The chromosomal location of PEA3 was also confirmed when the region surrounding the BRCA1 locus was sequenced (Friedman *et al.*,

1994). BRCA1, the breast/ovarian cancer susceptibility gene, is found at 17q21 (Hall *et al.*, 1990). Interestingly, HER-2 also maps to chromosome 17 at q12-21.32 (Popescu *et al.*, 1989).

CONCLUSIONS

Deletions of the 5' end of the PEA3 gene were assayed for their ability to drive the expression of the luciferase reporter gene in the FM3A cell line. The region which gave rise to the maximal promoter activity was a region containing the sequences -3 to +676, relative to the transcription start site of the PEA3 gene. The human homologue of mouse PEA3 and its putative promoter region were isolated from a genomic library and sequenced. Both the coding and the non-coding sequences surrounding the putative promoter were highly conserved between the mouse and human genes. The preservation of the consensus binding sites for regulatory factors and the functional analysis of the promoter deletion constructs suggested a role for sequences responsible for the expression of PEA3. It was also of interest to map the PEA3 gene to chromosome 17, a chromosome notorious for its involvement in human breast cancer.

REFERENCES

Alam, J. and J.L. Cook (1990) *Anal. Biochem.*, **188**, 245-254.

Alam, T., An, M.R. and J. Papaconstantinou (1992) *J. Biol. Chem.*, **267**, 5021-5024.

Baert, J-L., Monte, D., Musgrove, E.A., Albagli, O., Sutherland, R.L. and Y. de Launoit (1997) *Int. J. Cancer,* **70**, 590-597

Bailly, R.A., Bosselut, R., Zucman, J., Cormier, F., Delattre, O., Roussel, M., Thomas, G. and J. Ghysdael (1994) *Mol. Cell. Biol.*, **14**, 3230-3241.

Beaupain, D., Eleouet, J.F. and P.H. Romeo (1990) *Nucleic Acids Res.*, **18**, 6509-6515.

Ben-David, Y., Giddens, E.B., Letwin, K. and A. Bernstein (1991) *Genes Dev.*, **5**, 908-918.

Benz, C.C., O'Hagan, R.C., Richter, B., Scott, G.K., Chang, C-H., Xiong, X., Chew, K.,Ljung, B-M., Edgerton, S., Thor, A., and J.A. Hassell, *Oncogene,* in press.

Bosher, J.M., Totty, N.F., Hsuan, J.J., Williams, T. and H.C. Hurst (1996) *Oncogene,* **13**, 1701-1707.

Boulukos, K.E., Pognonec, P. Begue, A., Galibert, F. Gesquiere, J.D., Stehelin, D. and J. Ghysdael (1988) *EMBO J.*, **7**, 697-705.

Boyle, A.L., Feltquite, D.M., Dracopoli, N.C., Housman, D.E. and D.C. Ward (1992) *Genomics,* **12, *106-115.***

Brenner, S., Elgar, G., Sandford, R., Macrae, A., Venkatesh, B. and S. Aparicio (1993) *Nature,* **366**, 265-268.

Brown, L. and R. Baer (1994) *Mol. Cell. Biol.*, **14**, 1245-1255.

Brown, T.A. and S.L. McKnight (1992) *Genes Dev.*, **6**, 2502-2512.

Bucher, P. (1990) *J. Mol. Biol.*, **212**, 563-578.

Buijs, A., Sherr, S., van Baal, S., van Bezouw, S., van der Plas, D., Geurts van Kessel, A., Riegman, P., Lekanne Depres, R., Zwarthoff, E., Hagemeijer, A. and G. Grosveld (1995) *Oncogene*, **10**, 1511-1519.

Burtowski, S., Hahn, S. Guarent, L. and P.A. Sharp (1989) *Cell*, **56**, 549-561.

Chang, C-H., Scott, G.K., Kuo, W-L., Xiong, X., Suzdaltseva, Y., Park, J.W. Sayre, P., Erny, K., Collins, C., Gray, J.W. and C.C. Benz (1997) *Oncogene*, **14**, 1617-1622.

Chen, H., Ray-Gallet, D., Zhang, P., Hetherington, C.J., Gonzalez, D.A., Zhang D-E., Moreau-Gachelin, F. and D.G. Tenen (1995) *Oncogene*, **11**, 1549-1560.

Chen, J.H., Vercamer, C., Li, Z., Paulin, D., Vandenbunder, B. and D. Stehelin (1996) *Oncogene*, **13**, 1667-1675.

Crepieux, P., Leprince, D., Flourens, A., Albagli, O., Ferreira, E. and D. Stehelin (1993) *Gene Expression*, **3**, 215-225.

Dalton, S. & R. Treisman (1992) *Cell*, **68**, 597-612.

deWet, J.R., Wood, K., DeLuca, M., Helinski, D.R., and S. Subramani. (1985) *Mol. Cell. Biol.*, 7, 725-737.

Donaldson, L.W., Petersen, J.M. Graves, BJ., and L.P. McIntosh (1996) *EMBO J.*, **15**, 125-134.

Du, H. Roy, A.L. and R.G. Roeder (1993) *EMBO J.*, **12**, 501-511.

Feinberg, A.P. and B. Volgelstein (1983) *Anal. Biochem.*, **132**, 6-13.

Ferreri, K., Gill, G. and M. Montminy (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 1210-1213.

Friedman, L.S., Ostermeyer, E.A., Lynch, E.D., Welcsh, P., Szabo, C.I., Meza, J.E., Anderson, L.E., Dowd, P., Lee, M.K., Rowell, S.E., Ellison, J., Boyd, J. and M-C King (1995) *Genomics*, **25**, 256-263.

Funahashi, J. Kamachi, Y., Goto, K. and H. Kondoh (1991) *Nucleic Acids Res.*, **19**, 3543-3547.

Goebl, M.G., Moreau-Gachelin, F., Ray, D., Tambourin, P., Tavitian, A., Klemsz, M.J., McKercher, S.R. Celada, A., Van Bervern D. and A. Maki (1990) *Cell*, **61**, 1165-1166.

Golub, T.R., Barker, G.F., Bohlander, S.K., Hievbert, S.W., Ward, D.C., Bray-Ward, P., Morgan, E., Raimondi, S.C., Rowley, J.D. and D.G. Gilliland (1995) *Proc. Natl. Acad. Sci. USA,* **92**, 4917-4921.

Golub, T.R., Barker, G.F., Lovett, M. and D.G. Gilliland (1994) *Cell,* **77**, 307-316.

Gorman, C. (1985) *in* DNA Cloning II - A Practical Approach (Glover, D.M., Ed.), pp. 143-190, IRL Press, Oxford.

Grounds, M.D., Garrett, K.L., Lai, M.C., Wright, W.E. and M.W. Beilharz (1992) *Cell Tissue Res.,***267**, 99-104.

Hall, J.M., Lee, M.K., Newman, B., Morrow, J.E., Anderson, L.A., Huey, B. and M-C King (1990) *Science,* **250**, 1684-1689.

Heng, H. and L-C. Tsui (1993) *Chromasoma* **102**, 325-332.

Henthorn, P., Zervos, P., Raducha, M., Harris, H. and T. Kadesch, (1988) *Proc. Natl. Acad. Sci. USA,* **85**, 6342-6346.

Higashino, F., Yoshida, K., Fujinaga, Y., Kamio, K. and K. Fujinaga (1993) *Nucleic Acids. Res.,* **21**, 547-553.

Higashino, F., Yoshida, K., Noumi, T., Seiki, M. and K. Fujinaga (1995) *Oncogene,* **10**, 1461-1463.

Hoffman, A., Sinn, E., Yamamoto, T., Wang, J., Roy, A., et al., (1990) *Nature,* **346**, 387-390.

Hollywood, D.P. and H.C. Hurst (1993) *EMBO J.,* **12**, 2369-2375.

Ichikawa, H., Shimizu, K., Hayashi, Y. and M. Ohki (1994) *Cancer Res.,* **54**, 2865-2868.

Inglehart, J.D., Kraus, M.H., Langton, B.C., Langton, B.C., Huper, G., Kerns, B.J. and J.R. Marks (1990) *Cancer Res.,* **50**, 6701-6707.

Isobe, M., Yamagishi, F., Yoshida, K., Higashino, F., and E. Fujinaga (1995) *Genomics,* **28**, 357-359.

Jain, V. K. and I.T. Magrath (1991) *Anal. Biochem.,* **199**, 119-124.

Janhnecht, R., and A. Nordheim (1993) *Biochemica et Biophysica Acta,* **1155**, 346-356.

Jeon, I-S., Davis, J.N., Braun, B.S., Sublett, J.E., Roussel, M.F., Denny, C.T. and D.N. Shapiro (1995) *Oncogene,* **10**, 1229-1234.

Kaneko, Y., Yoshida, K., Handa, M., Toyoda, Y., Nishihira, H., Tanaka, Y., Sasaki, Y., Ishida, S., Higashino, F. and K. Fujinaga (1996) *Genes Chomosom. Cancer,* **15**, 115-121.

Kaya, M., Yoshida, K., Higashino, F., Mitaka, T., Ishii, S. and K. Fujinaga (1996) *Oncogene,* **12**, 221-227.

Kim, C.G., Swendeman, S.L., Barnhart, K.M. and M. Sheffery (1990) *Mol. Cell. Biol.,* **10**, 5958-5966.

King, C.R., Swain, S.M., Porter, L. Steinberg, S.M., Lippman, M.E. and E.P. Gelmann (1989) *Cancer Res.,* **49**, 4185-4191.

Kistler, B., Pfisterer, P. and T. Wirth (1995) *Oncogene,* **11**, 1095-1106.

Klemsz, M.J., Maki, R.A., Papayannopoulou, T., Moore, J. and R. Hromas (1990) *J. Biol. Chem.,* **268**, 5769-5773.

Koop, B.F. and J.H. Nadeau (1996) *Proc. Natl. Acad. Sci. USA,* **93**, 1363-1365.

Kozak, M. (1991) *J. Biol. Chem.,* **266**, 19867-19870.

Kraus, M.H., Popescu, N.C., Amsbaugh, S.C., and C. Richter-King (1987) *EMBO J.,* **6**, 605-610.

LaMarco, K.L., Thompson, C.C., Byers, B.P. Walton, E.M. and S.L. McKnight (1991) *Science,* **253**, 789-792.

Lassar, A.B., Davis, R.L., Wright, W.E., Kadesch, T. Murre, C., Voronova, A., Baltimore, D. and H. Weintraub (1991) *Cell,* **66**, 305-315.

Leprince, D., Gegonne, A., Coll, J., de Taisne, C., Schneeberger, A., Lagrou, C. and D, Stehelin (1983) *Nature,* **306**, 395-397.

Lessnick, S.L., Braun, B.S., Denny, C.T. and W.A. May (1995) *Oncogene,* **10**, 423-431.

Lichter, (1990) *Science,* **247**, 64-69.

Liu, D., Pavlopoulos, E., Modi, W., Moschonas, N. and G. Mavrothalassitis (1997) *Oncogene,* **14**, 1445-1451.

Lopez, M., Oettgen, P., Akbarali, Y., Dendorfer, U. and T. Libermann (1994) *Mol. Cell. Biol.,* **14**, 3292-3309.
MacGregor, G.R., Nolan, G.P., Fiering, S., Roederer, M. and L.A. Herzenberg (1991) *in Gene Expression in Vivo* (Murray E.J., and Walker, J.M., Eds.) pp. 217-236, Humana Press, Clinton, NJ.

Maldonado, E., Ha, I., Cortes, P., Weis, L. and D. Reinberg (1990) *Mol. Cell. Biol.* **10**, 6335-6347.

Martin, M.E., Piette, J., Yaniv, M., Tang, W-J., and W.R. Folk (1988) *Proc. Natl. Acad. Sci. USA,* **85**, 5839-5843.

Matrisian, L.M. (1994) *Ann. N.Y. Acad. Sci. USA,* 42-50.

May, W.A., Lessnick, S.L., Braun, B.S., Klemsz, M., Lewis, B.C., Lunsford, L.B., Hromas, R. and C.T. Denny (1993) *Mol. Cell. Biol.* **13**, 7393-7398.

Means, A.L. and P.J Farnham (1990) *Mol. Cell. Biol.,* **10**, 653-661.

Mermod, N., Williams, T.J. and R. Tjian (1988) *Nature,* **332**, 557-561.

Molnar, A. and K. Georgopoulos (1994) *Mol. Cell. Biol.,* **14**, 8292, 8303.

Monté, D., Baer, J-L., Defossez, P-A., de Launoit, Y. and D. Stehelin (1994) *Oncogene,* **9**, 1397-1406.

Monté, D., Coutte, L., Baert J-L., Angeli, I., Stehelin, D. and Y. de Launoit (1995) *Oncogene,* **11**, 771-779.

Montgomery, D.C. (1991) *in* Design and Analysis of Experiments. 3rd Edition. John Wiley and Sons, New York, USA.

Moreau-Gachelin, F., Tavitian, A. and P. Tambourin (1988) *Nature,* **331**, 277-280.

Nasrin, N., Buggs, C., Kong, X,F., Carnazza, J., Goebl, M. and M. Alexander-Bridges (1991) *Nature* **354**, 317-320.

Nerlov, C., De Cesare, D., Pergola, F., Caracciolo, A., Blasi, F., Johnsen, M. and P. Verde (1992) *EMBO J.,* **11**, 4573-4582.

Oettgen, P., Akbarali, Y., Boltax, J., Best, J., Kunsch, C. and T.A. Libermann (1996) *Mol.Cell. Biol.*, **16**, 5091-5106.

Ohno, T., Rao, V.N. and E.S.P. Reddy (1993) *Cancer Res.*, **53**, 5859-5863.

Oka, T., Rairkar, A. and J.H. Chen (1991) *Oncogene*, **6**, 2077-2083.

Papadopoulos, P., Ridge, S.A., Boucher, C.A., Stocking, C. and L.M. Wiedemann (1995) *Cancer Res.*, **55**, 34-38.

Parkes, H.C., Lillycrop, K., Howell, A. and Craig, R.K. (1990) *Br. J. Cancer*, **61**, 39-45.

Perrotti, D., Melotti, P., Skorski, T., Casella, I., Peschle, C. and B. Calabretta (1995) *Mol. Cell. Biol.*, **15**, 6075-6087.

Peter, M., Couturier, J., Pacquement, H., Michon, J., Thomas, G., Magdelnat, H and O. Delattre (1997) *Oncogene*, **14**, 1159-1164.

Peterson, M.G., Tanese, N., B.F. Pugh and R. Tjian 1990 *Science*, **248**, 1625-30.

Popescu, N.C., King, C.R. and M.H. Kraus (1989) *Genomics*, **4**, 362-366.

Prestridge, D.S. (1995) *J. Mol. Biol.*, **249**, 923-932.

Pugh, B.F. and R. Tjian (1991) *Genes Dev.*, **5**, 1935-1945.

Pugh, B.F. and R. Tjian (1992) *J. Biol. Chem.*, **267**, 679-682.

Quandt, K., Frech, K, Karas, H., Wingender, E. and Werner, T. (1995) *Nucleic Acid. Res.*, **23**, 4878-4884.

Rabindran, S.K., Giorgi, G., Clos, J. and C. Wu (1991) *Proc. Natl. Acad. Sci. USA* **88**, 6906-6910.

Rao, V.N, Huebner, K., Isobe, M., Ar-Rushdi, A., Croce, C.M. and E.S.P. Reddy (1989) *Science*, **244**, 66-70.

Rao, V.N., Papas, T.S. and E.S.P. Reddy (1987) *Science*, **237**, 635-639.

Ray, D ., Bosselut, R., Ghysdael, J., Mattei, M-G., Tavitian, A. and F. Moreau-Gachelin (1992) **12**, 4297-4304.

Reddy, E.S.P., Rao, V.N., and T.S. Papas (1987) *Proc. Natl. Acad. Sci.,* **84**, 6131-6135.

Romana, S.P., Mauchaffe, M., Le Coniat, M. Chumakov, I., Le Paslier, D., Berger, R. and O.A. Bernard (1995) *Blood,* **85**, 3662-3670.

Romana-Spica, V., Suziki, H., Georgiou, G., Chen, S-L., Ascione, R., Papas, T.S. and N.K. Bhat (1994) *Int. J. Oncology,* **4**, 521-531.

Rorth, P., Nerlov, C., Blasi, F., and M. Johnsen (1990) *Nucleic. Acids Res,* **18**, 5009-5017.

Sambrook, J., Fritsch, E.F. and T. Maniatus (1989) *in* Molecular Cloning-A Laboratory Manual. 2nd Edition. Cold Spring Harbour Laboratory Press, Cold Spring Harbour. New York.

Scott, G.K., Daniel, J.C., Xiong, X., Maki, R.A., Kabat, D. and C.C. Benz (1994) *J. Bio. Chem.,* **269**, 19848-19858.

Seto, E., Shi, Y. and T. Shenk (1991) *Nature* **354**, 241-245.

Singleton, T.P. and J.G. Strickler (1992) *Pathol. Annu.,* **1,** 165-190.

Skinner, A. and H.C. Hurst (1993) *Oncogene,* **8**, 3393-3401.

Slamon, D.J., Clark, G.M., Wong, S.G., Levin, W.J., Ullrich, A. and W.L. McGuire (1987) *Science,* **235**, 177-182.

Smale, S.T. and D. Baltimore (1989) *Cell,* **57**, 103-113.

Smale, S.T., Schmidt, M.C., Berk, A.J, and D. Baltimore (1990) *Proc. Natl. Acad. Sci. USA,* **87**, 4509-4513.

Smillie, D. (1993) MSc. Thesis, McMaster University, Hamilton, Ontario.

Sorensen, P.H.B., Lessnick, S.L., Lopez-Terrada, D., Liu, X.F., Triche, T.J. and C.T. Denny (1994) *Nat. Genetics* **6**, 146-151.

Southern, E (1975) *J. Mol. Biol.,* **98**, 503-517.

Struhl, K (1985) *Biotechniques,* **3**,452-453.

Taylor, J., Dupont-Versteegden, E.E., Davies, J.D., Hassell, J.A., Houle, J.D., Gurley, C.M. and C.A. Peterson, personal communication.

Thompson, C.B., Wang, C-Y, Ho, I-C., Bohjanen, P.R., Petryniak, B., June, C., Miesfeldt, S., Zhang, L., Nabel, G.J., Karpinski, B. and J.M. Leiden (1992) *Mol. Cell. Biol.*, **12**, 1043-1053.

Trimble, M.S., Xin, J-H., Guy, C.T., Muller, W.J. and J.A. Hassell (1993) *Oncogene,* **8**, 3037-3042.

Urano, F., Umezawa, A., Hong, W., Kikuchi, H. and J. Hata (1996) *Biochem. Biophys Res. Comm.*, **219**, 608-612.

Wasylyk, B., Hahn, S.L., and A. Giovane (1993) *Eur. J. Biochem.*, **211**, 7-18.

Watson, D.K., McWilliams, M.J., Lapis, P., Lautenberger, J.A., Schweinfest, C.W. and T.S. Papas (1988) *Proc. Natl. Acad. Sci.*, **85**, 7862-7866.

Weintraub, H., Tapscott, S.J., Davis, R.L., Thayer, M.J., Adam, M.A., Lassar A.B. and A.D. Miller (1989) *Proc. Natl. Acad. Sci. USA,* **86**, 5434-5438.

Wingender, E., Dietze, P., Karas, H. and R. Knuppel (1996) *Nucleic Acids Res.*, **24**, 238-241.

Wingender, E., Kel, A.E., Kel, O.V., Karas, H., Heinemeyer, T., Dietze, P, Knuppel, R., Romaschenko, A.G. and A. Kolchanov (1997) *Nucleic Acids. Res.*, **25**, 265-268.

Wu, T., Ikezono, T., Angus, C.W. and J.H. Shelhamer (1994) *Nucleic Acids Res.*, **22**, 5093-5098.

Xin., J-H., Cowie, A., Lachance, P. and J.A. Hassell (1992) *Genes Dev.* **6**, 481-496.

Zawel, L. and D. Reinberg (1993) *Prog. Nucleic Acid Res. Mol. Biol.* **44**, 67-108.

Zucman, J., Melot, T., Desmaze, C., Ghysdael, J., Peter, M., Zucker, J.M., Trich, T., Sheer, D., Turc-Carel, C., Ambros, P. Combaret, V., Lenoir, G., Aurias, A., Thomas G. and O. Dellatre (1993) *EMBO J.*, **12**, 4481-4487.

Appendix 1. Examples of Transcription Factor Consensus Matrices

Matrices were downloaded from the Transfac® database. The matrix contains the position in the consensus and a value is given for each nucleotide.

MATRIX table

Version: 2.5

DOCUMENTATION

AC M00032

ID....V CETS1P54_01

NA....c-Ets-1 p54

DT....EWI created ; 10.12.94.
DT....ewi updated ; 16.10.95.

PO	A	C	G	T
01	7	2	3	3
02	1	14	0	0
03	5	9	0	1
04	0	0	15	0
05	0	0	15	0
06	15	0	0	0
07	8	2	0	5
08	4	1	10	0
09	1	6	0	8
10	5	4	4	2

BF.... T00111; c-Ets-1 ; Species: mouse Mus musculus
BF.... T00114; c-Ets-1 54 ; Species: chick Gallus gallus

BA....15 selected binding sites for bacterially expressed murine factor

CC....Dissociation constants range between 0.038 nM and >3 nM

RF....References

TRANSFAC database
17-11-1995
transfac@gbf-braunschweig.de

TRANSFAC Browse Matrix MATRIX table

Version: 2.5

DOCUMENTATION

AC M00083

ID....V MZF1_01

NA....MZF1

DT....hiwi created ; 24.04.95.
DT....ewi updated ; 16.10.95.

PO	A	C	G	T
01	8	3	6	3
02	4	2	11	3
03	3	3	5	9
04	0	0	20	0
05	1	0	18	1
06	0	0	19	1
07	0	0	20	0
08	18	1	1	0

BF.... T00529; MZF-1 ; Species: human Homo sapiens

BA....20 selected binding sequences

CC....14-mers bound to bacterially expressed MZF1 zinc fingers 1-4

RF....References

TRANSFAC database
17-11-1995
transfac@gbf-braunschweig.de

Appendix 2. Genomic Human PEA3 Sequence

PEA3 sequence starting at three kilobasepairs upstream from the putative transcription start site is presented. Exon sequences are underlined.

GGGAGGTTGA GGTGTTAGGA TAGCTTGAAG CCAGGGAGTT TGAGGTTATA GTAAGCTATG ATGGCANCCA CTGTACTCCA ACCTGGGTGA CAGAGGGAGA CCCTCCTGTC

TCTCTAAAAA GAAAAAAAAA AAAAAAAAAA AAAAAAAAGG CTGGGCTTGG TGGCTCATGC CTGTAATCCC AGCACTTTAG GAAGCCAAGT GGATCACCTG AGGTCGGGAG

TTCGAGACCA GCCTGACCAA CATGGTGAAA CTCCGTCTGT ATTAAAAATA CAAAAAATTA GCCAGGTATG GTGGCGGGCA CCTGTAATCC CAGCTGCTCG GGAGGCTGAG

GCAGGAGAAT CGCTTGAACC CGGGAGGCGG AGGTTGCAGT GAGCCGAGAT TGCACCTTTG CACTCCAGTG TGGGCAACAA GAGTGAAACT CTGTCTCAAA AACAAAAACA

AAAAAACTCA GAACCAGAGG GGCACTGGCA TCCCTAGAGA GGTCCCATCC TACTGAAGTG CCCAGTACCA ACTTAAAGAC CTACTGGACT CTTCTTCGTT GTCATGGAGA

CCACATCAAA GCTGCCTTCC AGGAAGAGCC CTCTGTCCTC CTCCTCCACT TCTCTTGGCC AGAACTCACA CTTGTCTCTA AACTGGTTCC ATCCAGCCCA AAGCTGGGAT

TTGCTTGGGC TTGTTTCTCC TTCCTTGGGC AGGGTTCCTG CTAAGCTCAA TTAATTTGGT CTTTATTGTA CCTATTCACT ACACAAGCAT TTCATGCCAC CCTCATAAGC

ATGCTAACAG GCTGAGGAGG ACTTTCAGAG GCCAAATGAG CTCTGGAGAG CTCTCCATAG ATAAGGATGT CTATACCATG AGGCAGGATG AAAAACATCC ACACCAGAGA

TATAGTCAAG AATGTTGTAT ATGGCTAAAG GAACTGATTC AGGCTATCAT GGAAGGCTTC CTGGAGGAGA GGGACAATAG ACATGAGCCT TATAGAATAT TACAGCACCT

ATAGAAAGGG AAGCTTACTT CCATTTGCAC AAGCAGATCT CAATGGAATG AGCCCTTATG AACACTAAAC AGCAATTGTA TGTTTTAGCC TTAGTTCTGG CCAGACAGGC

CCAAAGAGCT TACATGTGTC ATTTCCTTCT CTGAAACTCC ACTGCCTCAT CTGTAAAGTG GCGGGTGTTA GGTTAGATGA ATTTGGGAAG CCACTCCTGC TGGCCTGGCA

TTTGAACATT CTTTTGTGTT TATCACAAAC AGGCAGGATA AAAAATCAGG CTGTCCTCAA AGGAGTGCAT AGGCTGGAGA CATAGAGAAT TGTCTCAGCA AAAAGGCCTT

TGAGGATTCA TGTCCTTTGT GCCCCCATGG CCACAGCCCA AGCACCCCAG GTCCCCTCAG ATGTCTGTCT TGGACATAGC TGCTTGCCTG AAACCCAGCA CTGTATCTGA

CACGATAAAT GTTTATTGAA AGAAGTGGTG AATGGATGAG ATGACAGAGA CTATGACCAG CTTCTTAGCT AAAACTGAAA CCCATCCTGC CTCCAGATCA TCAGAACCTT

TGGGGCTTTT GCTCCTTTTG CTGGCTCTCA GATAAGAGGG AAGATGGCTG AGAGGGAGCN ACACTTTTCG GGAAATCATT CTACCCCTGG GGGCTTCTGG TTTCTGCTGC

AATCAGTAGA AAGTGGGTCA CAAAACAGCT CTGCAGAGAA CAGACCCTTG AGGAAAAGCC TGGATTCTGT GCCAGTTTAC ATTTATTGAT CACCTACTGT GTATCAGTTG

TCACATAGCT TAACTCATTT AATCATGGCT TCAGAGGGGT GTACAGGAGA GTGTGAGGGA TTTGAGGCTG TAAACCCCAG AGGTGCTGAC TAGCTGAGTG ACCTCAGGAG

AGTTACCATC CCTCTCTGAG TTTCATTTTT AATATCTGTA AGATACAAAC GATCCATTTC GCCGGGTTAT GGGAACGAAA TGAGCACGTG GAAGCGGAGC GCTTCATAAA

TGCTAGTTAT TATTGTTACC TCTGGGCACA CAAGCCCGGT CGCCTCCCTC CTGGGCGCAC ACAGACATGT TCATACTCCC ACGTACACGC AGATGCAGAG ACGTGTGCAG

ACGCCAGCCG GAGGAGCACG CGGCGCATCG ACCCAAGACC GTGGCGAGAC GGTGAAGAAG CCGCGGGGTC CTCGGCTTCT CTCTTTTTCT CTCGGTTCTT TATGAATGAA

CTCCAGCCGC GAGTTTTATT TTTTTATGAA TGGGAGCCGG AGCACTGAAT GAAGCCGGGA GCCAACCCCT ACATTCTGAT TGGCCCCCGC GGCTGGGAAA CTCCGCCCCC

CGGCCAGGCC GGACGACTCT CCGGATTGGC GGGCGTGCCC AGCCCTGCTG GGGTTCCATT CACAAAAGTC AATGAAACAA AGGGAAACGG GGGGCGGTGT GGCGGGGCGG

GGGGTGGAGA GCTGAGGCCG AGCGAAGGAA ATGCACCAAT CAGCTGCTCC CCCGGGCTCA CAACTGTCTG CTGCGCCCGA AAAACAAGTC GGTACGNTGG GGACCCGGGG

CCGGGGCCGC CTTATTCCGG CATAGCCCCG CGGCCCTCGT TGCGGGCTCC AGGGCATGCT CGGGACCCCC CGCGGCTCCA GCCCAGACGC CCCGGCCTCA GGTGAGGCTG

CGGGAGGGAA AGGGAGGGAG AGTGTCTGTG ATTCCCCGCC TCGCCTCCAA TGCACACCAC ACTCACTCTC ACCACCCGCC TCCGCAGATT CCAGGGCGGG GGTGGGAGGT

GCTGGGATTC AGGGTTGGAG GGGGCGCGGG AGTTGTGGGC ATGCGATGGT GTGCGCTTCG CCTGTTCCTC CGCGCGCGCG CGCGTACACA CACACACACA CACACACACA

CACACACACA CACACACACG TTCTTATGTA ACCGAGCCCG GGTAAAGCAG GGCTGCAGAA AGCAGAAACG GCGAGCCCGG CTCCTGGGAG CAGGTGGGAC CTCCTTTGGT

CTTTGGCGGG AGTGGGGGTG GCGCTCGGCC TGGCAGGCGG AAACCTCGCC ATCGCCCCAG TGAGTCGGAG AAGTTTCGCC GTCGGACCCA GAGCCTTGGG ACTGCCCCCT

EXON 2

CCCCGCCCCC ATCGCAGCCC TCCCTGGACG GTGTGCGAAC GCAGCCCCCT CCCCCACCAG GTCTCGGCCC CCGCTTGGGG CCCCGGCCGT GCGGCCGGAG GGAGCGGCCG

GATGGAGCGG AGGATGAAAG CCGGATACTT GGACCAGCAA GTGCCCTACA CCTTCAGCAG CGTGAGCGCC GCGCCGGCCT CCACGCCCGC CCCGCCCCGC ACCCAGCCCC

EXON 3

TACTCTCACC ACAGCCCCCC TCCCCGCAGT CCCAGCGGAG TCCTCCCCCA NACTAAGCGC CACAAGGTGA CTCGCGGGCA TTCTCCCCTC TTCTAGNAGA AATCGCCCGG

AAATGGGAGC TTGCGCGAAG CGCTGATCGG CCCGCTGGGG AAGCTCATGG ACCCGGCTCC CTGCCGCCCC TCGACTCTGA AGGTAAAGAG CAAGAGAGAG AAACCGCGAC

CCTCCTCCGC CACACACACA CACACACACA CACACACACA CACACACACA CGGCCGATTC TGGAACCCCC GCCTGTTTCC CCAAAAAACT CGCCCTCCAA GGGATTCTGG

TCATTCTTCT CGGTCCCTAC TCCCAAGGAG GGGAGTCACC GGGTCTTGCT AGGGCCCCTT CTCTCTTACC GTCCCCTCCA CCTCACCTCC AATTCCACTA TTTACCCCTC

EXON 4

AGATCTCTTC CAGGATCTAA GTCACTTCCA GGAGACGTGG CTCGCTGAAG CTGAGTAGTT TTGTGACCTT TTCTATTTTG GGAGGCTGGA CAGGGGGACA CTCCGTAGCC

CCTGGTATTC CACCCTAAGC CCTACACTTC TCCTCTCCCT TCCCTGCCTC AAGGAGTTCT CCAAGGGCCC TCTCCCCAGG TCCCTTGCCA CTCTACTCAC CCCAGCATCC

AGCGTGGACA CTTATATNCG ATATTTATTT TCTCTACCCT GAGCCAGGCT TGTGTAACTA GGCTTGAGGA TGAATCAGAC CATTGTGTGG GGTATTTGCT TATTTGAATG

GAACTTGGCC CCAGGCCCAG GCCCATTTTC CTGGTTGAAA TTTTCATTTG TAAAGTTTGT TTTGCAGAAG CTGCCAGGCC AAGGTGCATG TTGGGGTGTT GTTGGTGGGG

GGTGGGGGAG AAAATNAATT TGTCACTCAT CCTCCAGAGT TGCCCAAGCA TTGATTCTTA TCCCCTGGGT TTCTCTTCTT CCCAACCCCC AGCACCAAAC AATTTACTGT

CCAAGCCAAA AGAACAACAA AAGTTTAAAT AATTCTTTTC GGCAGTCCTT TGGCACAGAG AGCCCAGGGC TCCCTCCCCA AATAGCCTTG ATTGTAATTC TAATAGCAAG

TGTTCCTCCC TCTTCCATCC CGCTTGAATA CCTCTGCCTC CAGGCAAAAT TCCCCCCTTC CTGTGTGTGC TGGGGAGTTA GCTGTGAGCT GGCTGGGGTG GGGTCCGTGG

CTGAGGTCCG TGGCTGGGTC TGATNGGGAA GTGGTGTGTT CAGGAGGTTT CCTGGTTGGA TTCTTCAGAT GCCTTTTGGT GGGGCCTGAG GGTGTCGCTG GTGGGGCTGC

AGGTGGGAGG GAAAATGGTG GTTTACACAG TCTCTNAGAG GGAAGGGTTA ACTACCCCCA TCTTAGGGTT TATTTTAACA CTATGAAAAT GAAGAATCCN NNACACTAGG

AACTCCTTTT TTGAAGTCTT TTTNAGGTAT TTTCCTTAGG AATCTCCTCA TCCCTCCACC CTGGGTATCT ATTTTTNTTA AAGCTCNAAT CCAACTACCC TGGAACTCTC

CNTAGAAGCT CGGAGTAAAA ACAAAATGAC CCCTGACTAC ATCCTTGGAG TGAGACTAGG TCTGTCTTGC CCACTATGAT ATTCTAAAAA TATATATGTT GACAGAATGA

CTATTTAGCA TANAGCAAGC ACTCAAGACA TTTTGTGGAA TTAATGGCTG AATGGCTGTG TGAGTGCTGG AGGGATCTGG AGGTTCTGAG GGTATTGTGT GCATCATCCC

CGACCCTGCA AGTCCTGCAT GCCCAGCTCC AATTTACAGG CAGANCTTTG CCCCGTCTCC CCATCCTTCT TCTGAGGTTC TGGAGCTGGG CTAGAGCTGA AGAATGCAGA

GGTGTCAGTT TCCGTGGGCT CTCAAAGACT TCCTTTCCCA GTGCCTGATT TCTCTCTGGC TCACCGGGCT TCATTCCTTG CCTGTCAGCT GAGAATCCCC ACAGTGGGCA

GCTCCTAAGG CTGTGAACTA AGGGTGGGCT GANGTTGGCA TTTTGGCAGC AGCTGGTCCA TCTGTGCCTG GGTTTGAGAT GGGAGGGGGA NGAAAATGCT CTCATCTCTC

TACCTTGGAA GTCCCACCCC TTGGCCAGGG GTCTGGAAAN GGTACAAATG TCTGTCTGTG TCAGTGTCGG TGTGACANCA AGTGTGTGTG TGTTGAGAGA GAGAGAGAGA

GAGAGAGAGA GAGAGAGAGC GAGAGTGTGT GTGTGTTAGT TACATGCTGC TGCCCTGTAC ATTTAAATCC AATTGGTGGG GAAGT----- ---------- CTGCAACCTC

TGCTTCGTGG GTTGCAATTC TAGTGCCTCA GCCTTCCGAG TAGCTGGGAT TACAGGTGCC CACCGCCACG CCNGANTAAT TTTTTATTTT TTTAGTAGAG ATGGGAGTTT

CACTTTGTTG GCCNGGCTGG TCTCAAACTC CTGATNTCAA GTGATCCTCC CACCTCAGCC TCCCAAAGTG GTGGGATTAC AGGCATGAAC CACTGCGCCC AGCCACCCTA

GGNGATTTTA TTTATTATTT AATTTTATTT TTCTTNGGAG ATGGAGTTTC ACTCTTGTTG CCCAGGCTGG AGTGCAGTGG TGCGATCTCA GCTCACTGCA ACCTCTGCCT

CCCAGGTTCA AGCGATTCTT CTGTCTCAGC TCCGGAGTAG CTGGGATTAT AGGCTCCCGC CACCACGACC AGCTAAGTTT TGTACTTTTA GTAAAGATGG GGTTTCGTCA

TGTTGGCCAG GCTGGTCTCG AACTCCGGAC CTCAGGGTAT CCACCCGCCT TGACTTCCCA AAGTGCTGGA TTATAGGCGT GAGCCACCAT GCCTGGCCGC CTCCTATCTG

ATTTTTAAAT CCACAGCAGC CCACTGCTTA TAATGCCTGC CCATCCTAGA TATGACTGTT CTGAGCTTTA AGGGAGTTTT GTAAANTCTG AGGGGCGAGG CTGTTTCACT

TCAGAAAAAT GGAGGAAATA AAGACAAAGC ATNTAGAAGA AGAAAGACTG GAATCTGCTC AGACAGGTTC TGAGAACAGT CTTCCCTCCT GGGTCTTAGT GGAGCCCAGA

GGTGGAGGAC ACTAGGACCT AGGGTGGAGC AGCCTTTAGT CTNCAAGCAA GCTTGTCCAA CCCGTGGGCC TGTGGGCCAC TTTCAGCCCA GGGTAGCTTT GAATGCTTCC

CAGGCCCAAC ATAAATTCCT TGTAAAATTT CTTAAAACAT TATGAGATTT GTTTGTGATT TTAAGTTCAT CAGCTATCGT TAGTGTTAGT GTATTTGTGT ATTTTATGTG

TGGCCCAAGA CAATTCTTAC AATGTGGACA CCCCTGATCT AAGGCTGTTG GTCCTTTCTT GGCTGTTGCT AATCTGCAGG GCTGTAATAA TGGAGCCTCC CAGGGGGACA

EXON 5

ACCAGGGCTC CGCCCCAAAC TCACCTCTGA CTTCTCTCTG CAGCTCAGGT ACCAGACAGT GATGAGCAGT TTGTTCCTGA TTTCCATTCA GAAAACCGTG AGTAGAGGTC

CCTGGGGAGA GGGTATAGGG AGTGGAGGAT GGCAGGAAGT GAGTCATAGA TAGAGAATCA CAAGAGAGCA GCTTGTTCTC TCCTAAAAAA CAGTGGGGTT TGATCTTAAA

GACCAAGGCA ATAACTGAAT GAAAACCTTG GTTGAAATCT GGGGCTTGAC ATGTTCTTGC TGTGTGACCT TAGAGCAGAC ACTATCCCTC TGGTTCTCCA TGGCCTTATA

TCTGAAGTGA GGGGCCTGGG CCAGTTCAGC AGAGTCCAGT TCTAGTGTTG TTTGTTTGTT TGTTTGTTTC CCCCCAAGAC AGATTCTTGC TCTGTCACCC AGGCTGGAGT

GCAGTGGCAC AATCTCGGCT CACTGCAACT TCCGCCTCCC GGGTTTAAGC CATTCTCCTG CCTCAGCTTC CANAGTAACT GGGATTACAG GCATGCACAA CCACACCCAG

CTAATTTTTG TATTTTTAGT AAANATGGGG TTTCACCATT TTGGCCAGGA TGGTCTCGAA TTCCTGACCT TGTGATCCGC CCACCTCGGC CTCCCATAGT GTTTGGACTA

CAGGCGTGAG CCCACTGCGC CCGCCCCCAC TTCTANTATT CTAAGTCACT TCCATAAACT GGGGACAATT TTGCCTCCCC AAGAGACATT TGGTAATATC TGTAGACATC

TTTGGTTGTC ACACTTGAGG GTGGGGTGCT ACTGGTGTCT AATGGATAGA GAGGCCAGGG ATACTGCTAA AAAGCATCCT ATGATGCTCA GGACCGCTCC TTACAATATT

ATTAATAAAA CACTATCTGG CCCACAGGTC ANTACTGCCA CTGTTGAGAA ACCCTGCTTT AGGCTGGGTN ---------- ---------- ---------- NNNNNNGGGG

CTTCCCAAAG GGGTGGGANA ANAGGNNNGA CCNNCGNGCC CGGCCAAGAA GATTTTTTAT ATCTTCAAAT AAGAAAACAT GTCANTCCCC TGGAATTTTC CAGGCANTCA

GGAAAAATAC AGGCCCTAGG TTGGAGAAAT AAGAAAACAT GTCANTCCCC TGGAATTTTC CAGGCANTCA GGAAAAATAC AGGCCCTAGG TTGGAGGTGC TCATTCCTTG

AAAGTGTTTT NTGGGGCCTG GGGTTGCATT AAAAGTTTTT GTTTTTTGTT TCTTTNNGAG ACGGAGTCTC GTTNTGTCAC CCAGGTTGGA GTGCAGTGGC ANGATTTNGG

CTCANTGCAA TCCCCGCNTC NTGGGTTCAA GCAATTCCCT TGCCCCAGCT TCNTGAGTAG NTGGGAANAC AGGCATCCGC CACCATCCCC GGCTAATTTT TTTTTGTATT

TTTAGTAGAG ACAGGGTTTC ACCATGTTGG CCAGGCTGGT TTTGGACTCC TGCCTCCCAA AGTGCTGGGA TTACAGGTGT GAGCTACCGT GCCCAGCCTA AATGTTGAAC

TTTAGATGGT TTATTTAGTC AATAATCAGA CACTAGTTTG ATGTATAAAC TAGACATAAT GTAAATAGCC AAAATGTTCT TGTCTCACTG CGGGTCTCAG AAAATTGATA

CCCAGAGAGG GAAGGAGGCT GTACCAGTCA CCCAGTTGGT CNTCAAAGNT CAGAAGGTAT CAGAAGCCCT GCCTGCNTNT TGTGTGACAG GNTCTGGCCT TGTTNTGTCC

EXON 6

CCTCCACCCC CACAGTAGCT TCCACAGCCC CACCACCAGG ATCAACAAGG AGCCCCAGAG TCCCCGCACA GACCGGCCCT GTCCTGCAGC AGGAAGCCGC CACTCCCCTA

CCACCATGGC GAGCAGTGCC TTTACTCCAG GTGAGCCCCA GCCTGGACCT GGTGAAGGGT AGGAGAAGGG TTCCAGAGGA GGGAGGCAGG GGGCACCCAT GAGCAGAGGC

AGCAGCTGAG ATGCAAGTTA GACAACAGAA AGGATTTTCT AGCACTTGTA AGGAAGCAAT ACAGCTAAGG ATTTCCTCTG AATAGACGGA AAACGCAGCA GAGGGCCTCA

GGAATATATA CTGAGGCCAG AGCAGATTCT GTTTCCACCA AGAGGCCCAC TGTCCCACCT AAACCATTGC TCTGAGCAAC AAATCCCCCC TGTATGAATC CCAGGGGTTC

AAGACTAGCC TGGGCAACAG AGGGAGACTG TCTCTGAAAA AGAAAAGAAT TGGCTAGGGC TGAGTCTGTG GNTCCTGCCG TTCCCTTCTC AGCCCCCTCT ATCCTTCCTT

EXON 7

CTCCTGTAAT CCCTGTGACC TCGCTTTCCT AGGAATGAAG GGTTCAAAGA TGCTTTCCAG CTTCTCTGCC CCCTCACCCC TGTCTCCTTC TCAACTCCAG TGCCGACCCC

CCCAGACAAA TCGCCATCAA GTCCTCCCCT GCCCCTGGTG CCCTTGGACA GTCGCCCCTA CAGCCCTTTC CCCGGGCAGA GCAACGGAAT TTCCTGAGAT CCTCTGGCAC

CTCCCAGCCC CACCCTGGCC ATGGGTACCT CGGGGAACAT AGGTAAGGAA ACAGGGAAAC CCCTGCTCAT GGGCCCAGGN GGNGTTTTTT CAATTCGGGG TGTTTNTCCA

TAAAGACTTG AGCTGTGGGG TGGTAGAGGC AGGGGGCAGT GGAATAGAAA GGGAGAAATA GTGACACCCN AACCCCTTAC CCCCCATTCN AGAGCCCATT GTAAAAGCNT

EXON 8

AGGAGTAACA GGTTNGAGGA AGGATTAATT TAAAGTTNCT TNTCTTCCTT NTCCATCCAA CCACATCCAG CTCCGTCTTC CAGCAGCCCC TGGACATTTG CCACTCCTTC

ACATCTCAGG GAGGGGGCCG GGAACCCCTC CCAGCCCCCT ACCAACACCA GCTGTCGGAG CCCTGCCCAC CCTATCCCCA GCAGAGCTTT AAGCAAGAAT ACCATGATCC

CCTGTATGAA CAGGCGGGCC AGCCAGCCGT GGACCAGGGT GGGGTCAATG GGCACAGGTA CCCAGGGGCG GGGGTGGTGA TCAAACAGGA ACAGACGGAC TTCGCCTACG

ACTCAGGTAA NAAACTGGGG TGGGGCAGGG TGGCATGTNA TCAANTGTTC ANTTAAGCCC ANAATGTCTC CCNANTTTTT TTACATTGCT GTTTACCATT TCATACATCC

AACCCACCCA CCCACTTTTT TTCTTCATAA TTTCAAAAGA ATGTTATTTT TCTTTCTTTT TCTTTTTTTT TTTGAAATGG GGTCCTGCTC TGGCCCAGGC TAGAATGCAG

TGGCACGATC TTCTGCTCAC TGCAACCTCT GCCTACCGGG TTCAAGCGAA TCTTGTNCCT CAGCCTCCCC AATTNCCTGG GATTACAGTC TTCTGCTCAC TGCAACCTCT

GCCTACCCGG GTTCAAGCGA TTCTTGTGCC TCAGCCTCCC GAGTACCTGG GATTACAGGC NCCAGCCACC AGCCACCGTG CCCGGCTAAT TTTTGTATTT TTAGTAAAAA

TGGGGTTTCG CCATGCTGGC CAGGATGGTC TCGAACTCCT GACTTCATAT GATCCATCGC CTCNGCCTCC CAAAGTGCTG TGATTATAGA CATGAGCCAC CTCACCCGGA

CAAATTATTA CTTTTCAATG AGCAAACCTA TATTCAGATT AAAAGATAAG GGGGGCCCCC GGGGCATGGT GAGTCATCTG TAATCCCANC ACTTTANGAG GCCATGNNNN

---------- ---------- NNNNCTTATG GAATGGAATG AAAGGAAGGG ATGTGTTTGT TAGACGTGAA AAACAGGGCC NAGGGCCCAN TGNTTTTTCC CCAGGGTGCA

TGCAGGGNAG GTTTTGTTTT TNGTCACACA AAAACACACA GGGGGGGGNG TGNACAGGGT TGGGAACCTG ACTTGNTGAC CGGACATTTT TGTCCATTNN GAGGGAGTNT

TAGGNGTGGA CACNGAATTC TTTCCCCGGA NGTTTATTTC ACCNTCCTTC NTCCCCNTGA ATTNGCAGGA AACAGNTGNT CCCNTGGCAG CTTGGGCCNT TNAGAACAGC

TTGCCCAGCC CCCGNTGCTG CTTTCCATGG CTTCAGCCGC AGCCNTCAAG TNGAGGAGGG GTTCCAGCAT CACACTCCTT NTGGGTGAAC TTTCCCCTGG GATTTTGTGG

TTGGCAGGCA ACCTGGGCAA AGAACAGTCA CCAGGAAGCA GGCTGGAAGG AAGAAATTNT TGAATGTGGA TAGGACTTCC TCNTCCCCTG CCCTCGAGCT CCACCCCAAG

EXON 9

CCATTCTCAC ATCACCCTTG TTCCCCCACA GATGTCACCG GGTGCGCATC AATGTACCTC CACACAGAGG GCTTCTCTGG GCCCTCTCCA GGTGACGGGG CCATGGGGTA

AGGCAGCCCC CCTTCCCCTG CCAAGCCCTC CATGTTGGGG GAGGGAGCGC TGCCATGGGG GAGGGTCTCC CTGGCCAGCA GTCCCTCTGG AGTCTCTGGG GTCTCATCGG

EXON 10

TGAACCCCCA GATCTGAGCA ACCCCCAATT TCCTCCACAA GGCTATGGCT ATGAGAAACC TCTGCGACCA TTCCCAGATG ATGTCTGCGT TGTCCCTGAG AAATTTGAAG

GTCAGAGAAG TGACTGTTGA TGGGAGGGTC AAGGTCTTAT CACGCTGTGT CCCTGCAGGA GTCACATCAC GTTTCATTGT TGCAAGAGGG TGGGAACCCA TAGAAAAGTA

EXON 11

CCTGGGGAAG ACCCCCTTCG AGAAATCCCT CGGGTGGGAG GCGGTTGCTT GGGGGGACCC AAAAAGACAC ANAACCCCAC ACCTCCTATT TTCGTGTTCC CCAGGAGACA

TCAAGCAGGA AGGGGTCGGT GCATTTCGAG AGGGGCCGCC CTACCAGCGC CGGGGTGCCC TGCAGCTGTG GCAATTTCTG GTGGCCTTGC TGGATGACCC AACAAATGCC

CATTTCATTG CCTGGACGGG CCGGGGAATG GAGTTCAAGC TCATTGAGCC TGAGGAGGTG GGCCTCTCGG ATGTTCCCAG CCCTCCTTTC CAAAGTTTAC AGCCTGGAGG

TGGGAGAACC TGGGAAATGG TGGCACGTGC CTCCATCATG ATTCTTGCTT TACCTAACCT GAATTCTTGC CAAGCCTAAG TCTGTGGGCT GATGCTTTGT TGCAGAGCTA

EXON 12

GCTTGGCACT TTGCACCAAG AATCTCAAGT GCTCTTTCTG ACCTTACCCC CATTTTTTCC TCTGCCAAGG TAGCATCTCT ACCCCAAAAC TGTTTTGTTC CCAGGTCGCC

AGGCTCTGGG GCATCCAGAA GAACCGGCCA GCCATGAATT ACGACAAGCT GAGCCGCTCG CTCCGATACT ATTATGAGAA AGGCATGCAG AAGGTGGGGG CTGTGGGTCT

AGGGACAAGG TGGTGGGGGG CAGTGGCTGT GAGAAGCTGA CTGGGGAGAG GGTCAGCAGG GCAGTTCTCA GCAACTTTGT GGGATCAAGA TAATGAATCA GATCCTGAAT

CAGTCAGAGA GNATAGGAAA TGTGGAGNAA ATCCCAAGGT TCTCCTCCCC CAAAAAAGTG CNACNCTGTA ACTGAGCAAG CCCAAGCATG GAGTAAGTTG AGATGAAAAA

GGAGTAAGAA CTGTGGAAGG GAGAGTCAGC TTCTCGGGAA CCAGCATGGG AGAGAAATGC CCCGAGCATC TGCCTGTNCT GGTAGAAGGG ACACATTCCC CACTCCCCCA

EXON 13

CCTNACCCCA TACAAAGGNT GGGNGGNTAG NCAGGCTGAN CAGACCTTCT CTCTCCCCAT AGGTGGCTGG TGAGCGTTAC GTGTACAAGT TTGTGTGTGA GCCCGAGGCC

CTCTTCTCTT TGGCCTTCCC GGACAATCAG CGTCCAGCTC TCAAGGCTGA GTTTGACCGG CCTGTCAGTG AGGAGGACAC AGTCCCTTTG TCCCACTTGG ATGAGAGCCC

CGCCTACCTC CCAGAGCTGG CTGGCCCCGC CCAGCCATTT GGCCCCAAGG GTGGCTACTC TTACTAGCCC CCAGCGGCTG TTCCCCCTCC CGCAGGTGGG TGCTGCCCTG

TGTACATATA AATGAATCTG GTGTTGGGGA AACCTTCATC TGAAACCCAC AGATGTCTCT GGGGCAGATC CCCACTGTCC TACCAGTTGC CCTAGCCCAG ACTCTGAGCT

GCTCACCCGA GTCATTGGGA AGGAAAAGTG GAGAAATGGC AAGTCTAGAA CTAGTGGATG CCCCCGGGCT GCAGGAATTC AGCTCAGCTT CTTCCTAGGT CCAAGCCCCC

CACACCTTTT CCCAACCACA GAGAACAAGA GTTTGTTCTG TTCTGGGGGA CAGAGAAGGC GCTTCCCAAC TTCATACTGG CAGGAGGGTG AGGAGGTNCA CTGAGCTCCC

CAGATCTCCC ACTGCGGGGA GACAGAAGCC TGGACTCTGC CCCACGCTGT GCCCTGGAGG GTCCGGTTTG TCAGTTCTTG GTGCTCTGTG TTCCCAGAGG CAGGCGGAGG

TTGAAGGAAA GGAACCTGGG ATGAGGGGTG CTGGGTATAA GCAGAGAGGA TGGGTTCCTG CTCCAAGGGA CCCTTTGCCT TTCTTCTGCC CTTTCCTAGG CCCAGGCCTG

GGTTTGTACT TCCACCTCCA CCACATCTGC CAGACCTTAA TAAAGGCCCC CACTTCTCCC CATNANATNT CCCCCTNTTC TCTCCTTTGC TNCCTCCTTN CNNAACACAC

C